1972

Aspects of functional iteration

Milton Eugene Winger
Iowa State University

72-26,953

WINGER, Milton Eugene, 1931-
ASPECTS OF FUNCTIONAL ITERATION.

Iowa State University, Ph.D., 1972
Statistics

University Microfilms, A XEROX Company, Ann Arbor, Michigan

Aspects of

functional iteration

by

Milton Eugene Winger

A Dissertation Submitted to the

Graduate Faculty in Partial Fulfillment of

The Requirements for the Degree of

DOCTOR OF PHILOSOPHY

Major Subject: Statistics

Approved:

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

For the Major Department

Signature was redacted for privacy.

For the Graduate College

Iowa State University
Ames, Iowa

1972

TABLE OF CONTENTS

I. INTRODUCTION

The topic of functional iteration dates back to the nineteenth century. It is generally accepted that the first significant treatises on the subject were those of Schroeder (1871) and Koenigs (1884). The topic also has current interest, due in part to the fact that extensive iterative computations can now be performed by machine.

Central to iteration processes is the notion of fixed point. In the classic works on iteration an "attractive" fixed point is used. That is, if a is a fixed point for f(x) and x > a, then $f(x) \leq f(f(x)) \leq \cdots \leq a$ while if x < a, then $f(x) \geq f(f(x)) \geq \cdots \geq a$, and, typically, when x is sufficiently close to a then successive self composition tends to draw the iterate towards a. Often, successive scalings of the functional argument yield convergence where it would not otherwise obtain, or modify the rate of convergence.

Most papers concerning iteration of functions treat iterates of analytic functions of a complex variable. In this discussion we shall confine ourselves to iteration of real transformations having domain and range in R_n, with some emphasis on R_1; for this latter case, classical results will be specialized to the reals to provide alternate derivations of limits of certain sequences of probability-related functions under iteration.

A well-known instance of functional iteration is the iterative composition of the probability generating functions in the theory of branching processes. We have included an introductory discussion of the Galton-Watson process in Section A of Chapter IV, dealing with the supercritical case.

A possibly new relation between probability generating functions and moment generating functions is developed, along with a possibly new continuous extension.

The connection between powers of a constant matrix (or scalar) and iterates of a transformation (or function) is emphasized in Section A of Chapter III by the notion of "easily iterated transformation". This is related to the concept "conjugate to a linear transformation" defined by Karlin and McGregor (1970), and the univariate "Schroeder iterates" of Szekeres (1958). Our definition pertains to real mappings having domain and range in R_n and employs a real $n \times n$ matrix, and a mapping and its left inverse map.

Perhaps the primary impetus for the author to initiate this line of research came from his exposure to the work of Thomas and David (1968), in the area of stochastic game theory. In that paper, a function called the maximin function was examined under iteration and scaling in a neighborhood of a fixed point. The fixed point in this case turns out to be a "repulsive" fixed point requiring suitable successive scalings of the functional argument to achieve convergence. Such scalings will be employed in most of our asymptotic limit derivations, which may be thought of as a "back door" approach compared with the classical results. The paper cited gives a proof of existence of a limit function for the sequence of scaled iterates of the maximin function and establishes continuity and monotonicity of the limit function. We establish in Sections C and D of Chapter IV that the limit function is analytic at zero and that its first derivative at zero is unity. Furthermore, a certain functional equation derived in that paper is shown to be a special form of Schroeder's equation when inverses exist, and the maximin function is shown to be "easily iterated" in a neighborhood of its fixed point.

Szekeres (1958), provides an extension to a real variable iteration theorem called the Koenigs-Kneser theorem. This theorem gives necessary and sufficient conditions for the existence of the inverse of the limit function when scaling a sequence of iterates, assuming an attractive fixed point. Our approach in the real variable theorem of Section C of Chapter IV, is the "backdoor" approach mentioned earlier. The function being iterated with argument successively scaled is the function inverse to that of Szekeres'. We exhibit necessary and sufficient conditions that the limit function exist and that its derivative be unity at zero.

By interpreting Koenigs' solution to Schroeder's equation in a real variable setting we prove in Section D of Chapter IV that a sequence of scaled inverse functions converges to an inverse limit function. The proof incorporates in an intermediary step the idea of convergence of a random variable in distribution.

We then give examples of various order statistic distribution functions to which the above discussion can be applied. Among these are the incomplete beta function and the distributions of the maximin, median and the k^{th} order statistic. Since convergence under scaling and iteration is a "central limit" type of result, it is reasonable to question existence of analogous "strong laws". We establish in Section A of Chapter V using bound functions, that the iterated median converges almost surely to 1/2, its fixed point on the interval (0,1). We further note that the iterated maximin and minimax also converge almost surely to their unique fixed points on (0,1). However, the method of proof does not work for an arbitrary order statistic. The extension of the method to other iterated order statistic distribution functions, using possibly other bound functions, is left unresolved.

In extending univariate results to the multivariate situation, we give a proof in Section B of Chapter III of the correspondence of successive powers of a matrix to successive iterates of a multivariate linear fractional mapping in R_n, which is a possibly new result.

The idea of using the inverse differential matrix for scaling purposes was briefly hinted at by Harris (1951), and treated by Karlin and McGregor (1970), without proof.

We suggest this approach, and, to be freed from the restriction of nonsingularity of the differential matrix, an "approximate inverse" is introduced in Section A of Chapter VI.

In Section A of Chapter VII we consider sufficient conditions for matrix scaling to yield limit maps analogous to those arising in the univariate case. A multivariate example of the supercritical multitype Galton-Watson process is discussed briefly in Section C of Chapter VII, the results being analogous to those of the univariate case.

Finally, in Chapter VIII, we develop bivariate analogs of the bound functions of the univariate case and formulate a definition of "bivariate bounding". However, the author has not found nontrivial maps all of whose iterates are bounded in the sense of this definition.

II. REVIEW OF MAP PROPERTIES

Let T be a mapping from R_n into R_n having domain D. If A is any subset of D, the image of A under T is $T(A) = \{T(x) : x \in A\}$. The range R of T is T(D). The inverse image of a subset B of R is $T^{-1}(B) = \{x : x \in D, T(x) \in B\}$.

T is said to be a continuous transformation if for every set B of R, open in the topology relative to R, $T^{-1}(B)$ is an open set in the topology relative to D.

If T* is a transformation from R onto D such that $T^*(T(x)) = x$ for $x \in D$, T* is a left inverse of T. Similarly, if T* is a transformation from R onto D such that $T(T^*(x)) = x$ for $x \in R$, T* is a right inverse of T. If T* is both a right and left inverse of T it is called an inverse of T, denoted T^-.

Theorem 2.1.1: If T has both a left inverse and a right inverse then these two inverses coincide.

Proof: Let T^+ be a right inverse and T* be a left inverse. Then $T^*(T(x)) = x$ for $x \in D$ and $T(T^+(x)) = x$ for $x \in R$. Hence we have $T^*(T(T^+(x))) = T^*(x)$ for $x \in R$. But $T^+(x) \in D$ and T*T is the identity map on D so we have $T^*(T(T^+(x))) = T^+(x)$ and $T^+(x) = T^*(x)$ for $x \in R$. □

The following corollary is an immediate consequence.

Corollary 2.1.1: All inverses coincide.

The following two definitions concern differentiable transformations.

Definition 2.1.1: A transformation T is differentiable at $x = a$ if there exists a linear transformation $L_a(h)$ such that

$$\lim_{h \to o} \frac{1}{||h||}[T(a + h) - T(a) - L_a(h)] = 0$$

where $||h||$ is the Euclidean vector norm.

Definition 2.1.2: If T is differentiable at $x = a$, the matrix corresponding to $L_a(h)$ in Definition 2.1.1 is called the differential matrix of T at a, denoted by $T'(a)$, and its determinant $|T'(a)|$ is called the Jacobian of T at A.

Useful criteria for establishing differentiability of T and obtaining $T'(a)$ are given in the following result, as found in Fleming (1965), page 101.

Theorem 2.1.2: A transformation is differentiable at a if and only if each of its components is differentiable at a. Furthermore, $T'(a)$ is the matrix of partial derivatives of the components of T evaluated at a.

Definition 2.1.3: Let $C_n(E)$ be the class of functions from R_n into R_1 such that the n^{th} partial derivatives exist and are continuous in an open region E of R_n. A mapping T is of class $C_n(E)$ if each component of T is of class $C_n(E)$.

The next two theorems concern inverse mappings and may be found in Buck (1965), pages 277-278.

Theorem 2.1.3: If T is a transformation with domain $D \subset R_n$ and range $R \subset R_n$ where T is of class $C_1(N)$, N a neighborhood of $x = a$, and

$|T'(a)| \neq 0$, then T is one-to-one on a neighborhood of a and has there an inverse T^-.

The inverse of Theorem 2.1.3 is called a local inverse. If $|T'(x)| \neq 0$ for all $x \varepsilon D$, T^- is said to be a global inverse for T one-to-one.

Theorem 2.1.4: Let T be a transformation of class $C_1(E)$ with $|T'(x)| \neq 0$ on E and let T map E one-to-one onto a set T(E). Then T^- is of class $C_1(T(E))$ and $[T^{-}{}'(a)] = [T'(a)]^{-1}$.

Definition 2.1.4: L(x), the vector of componentwise tangent planes at x = a for a mapping T is given by $L(x) = T(a) + L_a(x - a)$ where L_a is as defined in Definition 2.1.1.

Definition 2.1.5: A differentiable mapping T(x) is componentwise convex at x = a if $T(x) \geq L(x)$ for all x and T(x) is componentwise concave if the inequality is reversed.

A composition of mappings is usually called a product mapping. Suppose the maps R, S and T are such that R has domain A and range B, S has domain B and range C and T has domain C and range D. Then $SR(\cdot) \equiv S(R(\cdot))$ is a mapping with domain A and range C, $TS(\cdot) \equiv T(S(\cdot))$ is a mapping with domain B and range D and $TSR(\cdot) \equiv T(S(R(\cdot)))$ is a mapping with domain A and range D. Since the right side of the last identity equals both $TS(R(\cdot))$ and $T(SR(\cdot))$, the product of mappings is associative.

The next two theorems concern the chain rule and the differential approximation as applied to transformations. Proofs may be found in

Buck (1965), pages 264-265.

Theorem 2.1.5: Let T be of class $C_1(E)$, E open, and let S be of class $C_1(F)$, F open in T(E). Then ST is of class $C_1(E)$ and if $a \varepsilon E$ and $b = T(a)$, $(ST)'(a) = S'(b)T'(a)$.

The following theorem explores the uniformity of the limit in Definition 2.1.1.

Theorem 2.1.6: Let T be of class $C_1(D)$, D open, and let E be a closed bounded subset of D. Then if $a \varepsilon E$

$T(a + h) = T(a) + T'(a) \cdot h + R_a(h)$ where

$$\lim_{h \to o} \frac{||R_a(h)||}{||h||} = 0 \text{ uniformly for } a \varepsilon E.$$

Product mappings ST of special interest to us are those involving maps S and T that coincide on some joint domain D. In this case we shall adopt the notation $T(T(x)) = T^{(2)}(x)$, $T(T^{(2)}(x)) = T^{(3)}(x), \cdots$ $T(T^{(n-1)}(x)) = T^{(n)}(x)$.

Definition 2.1.6: $T^{(n)}(x) = T(T^{(n-1)}(x))$ is called the n^{th} compositional iterate of T or simply the n^{th} iterate of T.

If it happens that the domain D and range R of T are the same, we may employ the associativity property of product mappings and obtain the result $T^{(n+k)}(x) = T^{(n)}(T^{(k)}(x))$, $x \varepsilon D$, for all positive integers n and k. Under these circumstances, if T^{-} exists, we may include all integers with the understanding that $T^{(0)}(x)$ is the identity map on D and $T^{-n} = T^{-(n)}$.

A final remark regarding domain and range is prefaced by observing the function $T(x) = \frac{3x - 1}{5x - 2}$, where $T(T(x)) = T^{(2)}(x) = \frac{4x - 1}{5x - 1}$. Note that the exceptional set for $T(x)$ is $x = 2/5$ while the exceptional set for $T^{(2)}(x)$ is $x = 1/5$. However, thinking of the iterative scheme $T(T(x))$ we have the exceptional set $x = 2/5$ and $1/5$.

This illustrates the fact that when $T(x)$ maps D onto R where $D \neq R$, typically the iterate $T^{(n)}(x) = T(T(\cdots(T(x))\cdots))$ is well-defined only on a subset of D which depends on the interplay of exceptional sets acquired at each stage of iteration.

III. EASILY ITERATED MAPS AND EXAMPLES

A. Definitions and Basic Concepts

It will be of consequence for us to characterize a class of transformations that retain much of their original form in some sense under iteration. This will facilitate expressing the n^{th} compositional iterate of a mapping in a closed form and render it useful for further analysis.

In the classical works of Schroeder (1871) and Koenigs (1884), several functional equations assumed importance.

$$\emptyset(f(z)) = \beta\emptyset(z)$$

is known as Schroeder's equation where β, z, f and $\emptyset$ are complex.

$$\emptyset(g(z)) = f(\emptyset(z)) \tag{3.1.1}$$

is sometimes called Koenigs' equation and we shall encounter it in a real variable setting later on.

We shall make use of the classical solutions to these equations to obtain certain probabilistic results in Chapter IV.

If in Equation (3.1.1) $\emptyset$, f, g and z are taken as real and we postulate the existence of $\emptyset^{-}$, we may set $x = \emptyset(y)$ so $y = \emptyset^{-}(x)$ giving the equation

$$f(x) = \emptyset(g(\emptyset^{-}(x))). \tag{3.1.2}$$

It is easily seen that $f(f(x)) = \emptyset(g(g(\emptyset^{-}(x))))$ or $f^{(2)}(x) = \emptyset(g^{(2)}(\emptyset^{-}(x)))$. Then according to Definition 2.1.6, $f^{(3)}(x) = f(f^{(2)}(x))$ so we have $f^{(3)}(x) = \emptyset(g(g^{(2)}(\emptyset^{-}(x)))) = \emptyset(g^{(3)}(\emptyset^{-}(x)))$. Proceeding in this manner $f^{(n)}(x) = \emptyset(g^{(n)}(\emptyset^{-}(x)))$ is the n^{th} iterate of f. Since $\emptyset^{-}$ is on R and $\emptyset$ is onto R, f always has coincident domain and range if g is from D onto D. If the domain and range of g are the domain of $\emptyset$, then $f^{(n)}$ is a map with domain and range equal to the range of $\emptyset$.

If we consider mappings from R_n into R_n we may obtain analogous results. Let $\emptyset^*$ be a left inverse of $\emptyset$. We may observe the simple lemma that follows.

Lemma 3.1.1: If $T(x)$ is a transformation from D onto D, D a subset of R_n, where T is of the form $T(x) = \emptyset F \emptyset^*(x)$ on D, then the n^{th} iterative composition map is $T^{(n)}(x) = \emptyset F^{(n)} \emptyset^*(x)$ on D.

Proof: $T^{(2)}(x) = T(T(x)) = \emptyset F \emptyset^* \emptyset F \emptyset^*(x) = \emptyset F^{(2)} \emptyset^*(x)$. By Definition 2.1.6, $T^{(3)}(x) = T(T^{(2)}(x))$ so we have $T^{(3)}(x) = \emptyset F \emptyset^* \emptyset F^{(2)} \emptyset^*(x) = \emptyset F^{(3)} \emptyset^*(x)$ and proceeding inductively we have the result. $\square$

We now make a basic definition which will be used frequently in which we think of F replaced by a matrix A in Lemma 3.1.1.

Definition 3.1.1: Let $\emptyset$ be a mapping from D onto R, both subsets of R_n, with left inverse $\emptyset^*$. Any mapping T of R onto itself that can be expressed in the form $T(x) = \emptyset A \emptyset^*(x)$, where A is an $n \times n$ real matrix such that $A(D) \subset D$, shall be said to be an easily iterated mapping on R, denoted e.i. mapping.

(Karlin and McGregor (1970), used the term "conjugate to a linear transformation" as a related concept.)

Theorem 3.1.1: If $\emptyset$ is a mapping from D onto R and F(x) is an e.i. map on D, then $T(x) = \emptyset F \emptyset^*(x)$ is an e.i. map on R.

Proof: If F(x) is e.i. on D, then for some map G, $F(x) = GAG^*(x)$ on D where A is a real $n \times n$ matrix such that $A(R) \subset R$ and $G^*G(x) = x$ on R.

Then $T(x) = \emptyset GAG^*\emptyset^*(x)$ on R but $G^*\emptyset^* = (\emptyset G)^*$ so $T(x) = (\emptyset G)A(\emptyset G)^*(x)$. Letting $\emptyset G = H$, $T(x) = HAH^*(x)$ where H is from R onto R and $A(R) \subset R$. $\square$

As a partial converse we have the following theorem.

Theorem 3.1.2: If $T(x) = \emptyset F\emptyset^*(x)$ is an e.i. map where $\emptyset^*$ is also a right inverse, then F is an e.i. map.

Proof: Since $T(x)$ is e.i. we may write $T(x) = GAG^*(x) = \emptyset F\emptyset^*(x)$. Let $x = \emptyset(y)$ so $y = \emptyset^*(x)$. Then since $\emptyset F(y) = GAG^*(y)$ and taking $\emptyset^*$ of both sides gives $F(y) = (\emptyset^*G)A(G^*\emptyset)(y)$ where $(G^*\emptyset)$ is a left inverse of $(\emptyset^*G)$ we have the result. $\square$

If $T(p) = p$, p is said to be a fixed point for T. Since this concept is of major concern in iteration analysis the following theorem is useful.

Theorem 3.1.3: If $T(x) = \emptyset A\emptyset^*(x)$ where $T(p) = p$ and $A - I$ is non-singular, then $\emptyset^*(p) = 0$.

Proof: Since $T(p) = p$, $\emptyset A\emptyset^*(p) = p$ and taking $\emptyset^*$ of both sides gives $A\emptyset^*(p) = \emptyset^*(p)$. Then $(A - I)\emptyset^*(p) = 0$ so $\emptyset^*(p) = 0$. $\square$

In R_1, if $\emptyset'(0)$ exists and $\emptyset'(0) \neq 0$ where $T(x) = \emptyset(a\emptyset^*(x))$, then if $\emptyset^*$ is an inverse at $\emptyset(0)$ the chain rule yields

$$T'(p) = \emptyset'(a\emptyset^*(p)) \cdot a \cdot \emptyset^{*\prime}(p).$$

By the previous theorem, if $a > 1$ we would have

$$T'(p) = \emptyset'(0) \cdot a \cdot \emptyset^{*\prime}(p) = \emptyset'(0) \cdot a \cdot \frac{1}{\emptyset'(0)} = a.$$

This is, in fact, a special case of the following theorem.

Theorem 3.1.4: If $T(x) = \emptyset A \emptyset^{-}(x)$ where $\emptyset$ and $\emptyset^{-}$ are differentiable at the required points and $T(p) = p$, then A and $T'(p)$ are similar matrices provided $A - I$ and $\emptyset'(0)$ are nonsingular.

Proof: Let $\emptyset^{-}(x) = G(x)$. By the chain rule for map composition we have $T'(x) = \emptyset'(AG(x)) \cdot A \cdot G'(x)$. By Theorem 3.1.3, $G(p) = 0$ so $T'(p) = \emptyset'(0) \cdot A \cdot G'(p)$ but $G'(p) = [\emptyset'(G(p))]^{-1} = [\emptyset'(0)]^{-1}$ so $T'(p) = [\emptyset'(0)]A[\emptyset'(0)]^{-1}$. By definition of similar matrices, $T'(p)$ is similar to A. □

The following two corollaries are immediate consequences.

Corollary 3.1.1: $|A|$ is the Jacobian of T at p, $|T'(p)|$.

Corollary 3.1.2: If under the conditions of Theorem 3.1.4, A and $\emptyset'(0)$ commute, then $T'(p) = A$.

We shall employ several terms concerning easily iterated mappings which will be explained by the following definitions.

Definition 3.1.2: The set of easily iterated mappings corresponding to a parametric family of mappings $\{\emptyset\}$ each of which possesses a left inverse $\emptyset^*$ shall be called an easily iterated family of mappings.

Definition 3.1.3: If two easily iterated mappings T_1 and T_2 differ with respect to A but not with respect to $\emptyset$ they will be said to be of identical form.

<u>Definition 3.1.4</u>: If T_1 and T_2 are easily iterated mappings belonging to the same family they will be said to be of almost identical form.

<u>Definition 3.1.5</u>: If T_1 and T_2 differ with respect to $\emptyset$ but not with respect to A they will be said to be of similar form.

(The terminology for the last definition is suggested by the idea of similar matrices, i.e., PAP^{-1} and QAQ^{-1} are similar matrices.)

It will be instructive to think of an easily iterated map as composed of a core matrix A and an edge mapping $\emptyset$ whose left inverse exists. In this context, e.i. maps of similar form have the same cores. Maps of identical form have the same edges while maps of almost identical form would have edges belonging to the same parametric family.

<u>Definition 3.1.6</u>: If $T_1T_2(x) \equiv T_2T_1(x)$, T_1 and T_2 are said to be reversible..

<u>Theorem 3.1.5</u>: Two easily iterated maps of identical form are reversible if and only if their cores commute.

<u>Proof</u>: Let $T_1(x) = \emptyset A_1\emptyset^*(x)$ and $T_2(x) = \emptyset A_2\emptyset^*(x)$. If $A_1A_2 = A_2A_1$, then $T_1T_2(x) = \emptyset A_1A_2\emptyset^*(x) = \emptyset A_2A_1\emptyset^*(x) = T_2T_1(x)$. Reversing the steps yields the converse. $\square$

In R_1, the cores are scalar and commute so all e.i. functions of identical form commute.

Later we shall find it convenient to be able to translate mappings in such a way as to be able to treat the origin as the fixed point under consideration. The procedure is outlined by the following four steps:

(1) Let $T(x)$ be e.i. and let $p \neq 0$ be the fixed point under consideration.

(2) Let $S(x) = x - p$ and form the mapping $R(x) = STS^{-}(x) = T(x + p) - p$, so for $x = 0$, $R(0) = T(p) - p = p - p = 0$ and 0 is a fixed point for the e.i. mapping R.

(3) $R^{(n)}(x) = ST^{(n)}S^{-}(x) = T^{(n)}(x + p) - p$ shows that $R^{(n)}$ and $T^{(n)}$ are related in a simple manner.

(4) To return to $T(x)$ we use the relation $T^{(n)}(x) = S^{-}R^{(n)}S(x) = R^{(n)}(x - p) + p$.

To illustrate the ideas of this section we shall present some examples in the next section.

B. Examples

<u>Example 3.2.1</u>: Let $f(x) = x^a$ where $a > 0$ and $x > 0$. Taking $\emptyset(y) = \exp y$ and $\emptyset^{-}(y) = \ln y$ we have $f(x) = x^a = \exp(\ln x^a) = \exp(a \ln x)$. Then $f(x) = \emptyset(a(\emptyset^{-}(x))$ so $f^{(n)}(x) = \emptyset(a^n\emptyset^{-}(x))$.

To check, note that $f(f(x)) = (x^a)^a = x^{a^2}$, $f(f^{(2)}(x)) = (x^{a^2})^a = x^{a^3}$, $\ldots, f(f^{(n-1)}(x)) = (x^{a^{n-1}})^a = x^{a^n}$. Note that $\emptyset(y)$ is free of parameters so $f_1(x) = x^a$ and $f_2(x) = x^b$ would be of identical form for $a \neq b$, $b > 0$.

<u>Example 3.2.2</u>: Let $f(x) = 1 - (1 - x)^a$, $a > 0$, $0 < x < 1$. Then by taking $\emptyset(y) = 1 - \exp y$ and $\emptyset^{-}(y) = \ln(1 - y)$ we have

$f(x) = 1 - (1 - x)^a = 1 - \exp(\ln(1 - x)^a) = 1 - \exp(a \ln(1 - x)) = \emptyset(a\emptyset^-(x))$. Clearly we have f(x) easily iterated and $f^{(n)}(x) = \emptyset(a^n\emptyset^-(x))$.

To check this briefly, $f(f(x)) = 1 - \{1 - [1 - (1 - x)^a]\}^a = 1 - [(1 - x)^a]^a = 1 - (1 - x)^{a^2} = f^{(2)}(x)$. Proceeding similarly we see that $f^{(n)}(x) = 1 - (1 - x)^{a^n}$. Again, as in the previous example, $f_1(x) = 1 - (1 - x)^a$ and $f_2(x) = 1 - (1 - x)^b$ are of identical form for $b > 0$, $a \neq b$.

<u>Example 3.2.3</u>: Let $f(x) = \frac{ax}{bx + 1}$; $x \neq -\frac{1}{b}$, $a > 1$, $b \neq 0$.

If we set $\emptyset(y) = \dfrac{y}{\left(\frac{b}{a - 1}\right)y + 1}$ and $\emptyset^-(y) = \dfrac{y}{1 - \left(\frac{b}{a - 1}\right)y}$, f(x) may be written in the following form:

$$f(x) = \frac{ax}{bx + 1} = \frac{ax}{\left(\frac{b}{a - 1}\right)ax - \left(\frac{b}{a - 1}\right)x + 1}$$

$$= \frac{\left[\dfrac{ax}{1 - \left(\frac{b}{a - 1}\right)x}\right]}{\left(\frac{b}{a - 1}\right)\left[\dfrac{ax}{1 - \left(\frac{b}{a - 1}\right)x}\right] + 1} = \emptyset(a\emptyset^-(x)) \qquad (3.2.1)$$

At this point we remark that the domain of validity for $f(x) = \emptyset(a\emptyset^-(x))$ must be $D_1 = \{x : x \neq -\frac{1}{b}, x \neq \frac{a - 1}{b}\}$. However, to insure that the internal operations in taking iterates of (3.2.1) are well-defined we must further restrict the domain by excluding all exceptional sets. Define $W = \{-\frac{1}{b}, \frac{a - 1}{b}\}$ so $D_1 = \mathcal{R}_1 - W$. The domain needed to accommodate all iterates is $D = R_1 - W - f^{-1}(W) - f^{-1}(f^{-1}(W)) - \cdots$. This technicality will be mentioned further in the multivariate case. (The minus signs in the expression for D are to be construed in the set-theoretic sense.)

We may now write $f^{(n)}(x) = \emptyset(a^n\emptyset^-(x))$, $x \in D$.

This example is a particular case of a class of transformations called linear fractional transformations. They are useful in iterative analysis because of their ease of regeneration. In particular, all compositional iterates of linear fractional maps are themselves linear fractional.

In the univariate case, the general linear fractional function is of the form $f(x) = \frac{a_1x + a_2}{c_1x + c_2}$. Assuming proper restrictions on the domain, it has been observed that $f^{(n)}(x)$ may be obtained explicitly by observing the matrix expression

$$\begin{bmatrix} a_1 & a_2 \\ c_1 & c_2 \end{bmatrix}^n = \begin{bmatrix} a_1(n) & a_2(n) \\ c_1(n) & c_2(n) \end{bmatrix}$$ and setting up the correspondence

$$f^{(n)}(x) = \frac{a_1(n)x + a_2(n)}{c_1(n)x + c_2(n)}.$$

We shall generalize this to the multivariate case in the next theorem. A further comment concerning Example 3.2.3 is that $\emptyset$ depends on parameters a and b and so defines a family of easily iterated functions according to Definition 3.1.2.

<u>Theorem 3.2.1</u>: Let A be an $n \times (n + 1)$ real matrix, let c be a $1 \times (n + 1)$ row vector and let P be the $(n + 1) \times (n + 1)$ matrix $\begin{bmatrix} A \\ c \end{bmatrix}$. Let T be a mapping on R_n to R_n defined by

$$T : u_i = \left(\sum_{j=1}^{n} a_{ij}x_j + a_{i,n+1}\right)/\left(\sum_{j=1}^{n} c_j x_j + c_{n+1}\right);\ i = 1,2,\cdots,n.$$

Also, let T_P be a mapping on R_{n+1} to R_{n+1} defined by $T_P : Py/cy$ where $y = \begin{bmatrix} x \\ 1 \end{bmatrix}$. Then the iterate of T may be put into correspondence with the iterates of P by the left-hand and right-hand sides of the following relations:

$$\begin{bmatrix} T^{(n)}(x) \\ 1 \end{bmatrix} = T_P^{(n)} \begin{bmatrix} x \\ 1 \end{bmatrix} = \frac{P^n \begin{bmatrix} x \\ 1 \end{bmatrix}}{cP^{n-1} \begin{bmatrix} x \\ 1 \end{bmatrix}} = T_P\left(P^{n-1} \begin{bmatrix} x \\ 1 \end{bmatrix}\right). \qquad (3.2.2)$$

We assume, in analogy to prior remarks, the x-domain is such that these iterates can be computed.

<u>Proof</u>: Assuming the correct domain, we argue by induction that

$$T_P^{(n+1)} \begin{bmatrix} x \\ 1 \end{bmatrix} = T_P\left(T_P^{(n)} \begin{bmatrix} x \\ 1 \end{bmatrix}\right) = T_P \begin{bmatrix} T^{(n)}(x) \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} T(T^{(n)}(x)) \\ 1 \end{bmatrix} = \begin{bmatrix} T^{(n+1)}(x) \\ 1 \end{bmatrix},$$

and similarly for the other two equalities in (3.2.2). □

The next example provides a multivariate analog to Example 3.2.3.

<u>Example 3.2.4</u>: Let $T(x) = \dfrac{Ax}{cx + 1}$ be a mapping on R_n - E where E is the plane $cx + 1 = 0$, where A is $n \times n$ with $A - I$ nonsingular and c is $1 \times n$.

Set $\emptyset(w) = \dfrac{w}{c(A - I)^{-1}w + 1}$, $\emptyset^{-}(w) = \dfrac{w}{1 - c(A - I)^{-1}w}$ where w is an $n \times 1$ column vector. Then

$$T(x) = \frac{Ax}{cx + 1} = \frac{Ax}{c(A - I)^{-1}(A - I)x + 1} = \frac{Ax}{c(A - I)^{-1}x - c(A - I)^{-1}x + 1}$$

$$= \frac{\left(\dfrac{Ax}{1 - c(A - I)^{-1}x}\right)}{c(A - I)^{-1}\left(\dfrac{Ax}{1 - c(A - I)^{-1}x}\right) + 1} = \emptyset A \emptyset^{-}(x),$$

for $x \in R_n - E - W$, where W is the set of all x such that at least one of the steps involved in forming $\emptyset A \emptyset^{-}(x)$ is not well-defined.

Moreover, on the set $D = R_n - W - T^{-1}(W) - T^{-1}(T^{-1}(W)) - \cdots$ we have

$$T^{(n)}(x) = \emptyset A^n \emptyset^{-}(x) = \frac{A^n x}{c(A - I)^{-1}(A^n - I)x + 1}.$$

Note that this is a particular case of the transformation defined in Theorem 3.2.1 where $a_{i,n+1} = 0$ for $i = 1,2,\cdots,n$.

Example 3.2.5: Let T be the mapping whose domain and range are $\{(x,y) : x > 0, y > 0\}$ and T is defined by

$$T : \begin{array}{l} u = \emptyset_1[a_{11}\emptyset_1^{-}(x) + a_{12}\emptyset_2^{-}(y)] \\ \\ v = \emptyset_2[a_{21}\emptyset_1^{-}(x) + a_{22}\emptyset_2^{-}(y)] \end{array} . \tag{3.2.3}$$

We may write this more efficiently in the matrix operator form

$$T(w) = \begin{bmatrix} u \\ \\ v \end{bmatrix} = \begin{bmatrix} \emptyset_1 & 0 \\ \\ 0 & \emptyset_2 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} \\ \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} \emptyset_1^{-} & 0 \\ \\ 0 & \emptyset_2^{-} \end{bmatrix} \begin{bmatrix} x \\ \\ y \end{bmatrix} . \tag{3.2.4}$$

where the operators proceed from right to left. If $w = \begin{bmatrix} x \\ y \end{bmatrix}$, we have (3.2.4) in the form

$$T(w) = \emptyset A \emptyset^{-}(w), \text{ where } \emptyset = \begin{bmatrix} \emptyset_1 & 0 \\ 0 & \emptyset_2 \end{bmatrix},$$

so clearly $T(w)$ is easily iterated.

As a specific case, suppose we observe the bivariate analog of Example 3.2.1:

$$T: \begin{array}{l} u = x^{a_{11}} \cdot y^{a_{12}} = \exp(a_{11} \ln x + a_{12} \ln y) \\ v = x^{a_{21}} \cdot y^{a_{22}} = \exp(a_{21} \ln x + a_{22} \ln y) \end{array}$$

Here we have $\emptyset_1(t) = \emptyset_2(t) = \exp t$ and $\emptyset_1^{-}(t) = \emptyset_2^{-}(t) = \ln t$ so clearly T may be thought of in the operator form (3.2.4). Example 3.2.2 is also a special case of this "diagonal $\emptyset$" type of map.

There is no need to restrict the map in Example 3.2.5 to R_2 since the diagonal $\emptyset$ operator makes the n^{th} iterate dependent on the powers of A and the example may as well have been stated for R_k, $k \geq 2$.

<u>Example 3.2.6</u>: As final examples consider the univariate bounding functions employed by Thomas and David (1968), page 246.

$$\mu(v) = a\left(\frac{v}{a}\right)^b;\ 0 \leq v \leq 1,\ b > 1$$

$$\lambda(v) = 1 - (1 - a)\left(\frac{1 - v}{1 - a}\right)^b;\ 0 \leq v \leq 1,\ b > 1.$$

Example 3.2.1 is a special case of $\mu(v)$ with $a = 1$ and Example 3.2.2 is a special case of $\lambda(v)$ with $a = 0$.

$\mu(v) = \emptyset(b\emptyset^{-}(v))$ where $\emptyset(t) = ae^t$ and

$\emptyset^{-}(t) = \ln\left(\frac{t}{a}\right)$. Hence, $\mu(v)$ is e.i. and $\mu^{(n)}(v) = a\left(\frac{v}{a}\right)^{b^n}$.

$\lambda(v) = \emptyset(b\emptyset^{-}(v))$ where $\emptyset(t) = 1 - (1 - a)e^{t}$ and

$\emptyset^{-}(t) = \ln\left(\frac{1 - t}{1 - a}\right)$. From this we are able to write

$\lambda^{(n)}(v) = 1 - (1 - a)\left(\frac{1 - v}{1 - a}\right)^{b^n}$.

The functions λ and μ were used to bound ϕ, the maximin function, on the interval $[0,1]$ and characterize it, in a sense, near a fixed point on $(0,1)$. We shall discuss the iterated maximin and other functions in the last section of Chapter IV and show they are easily iterated. The bound functions λ and μ will be employed in Chapter V in connection with almost sure convergence proofs.

IV. ASYMPTOTIC ASPECTS OF SCALING AND ITERATION OF UNIVARIATE FUNCTIONS

A. Scaled Iterations of Probability Generating Functions in Branching Processes

In this section we shall explore certain limit distributions obtained from iteration and scaling of probability generating functions associated with the basic Galton-Watson type branching process. The cascade or branching processes offer a good example of some of the iterative limits to be analyzed in detail later in this chapter. We shall confine the discussion to the "supercritical" case, i.e. where $\mu > 1$. Elementary discussions of generating functions and branching processes may be found in Feller (1968), pages 264-267 and 293-300. A more advanced discussion is found in Harris (1963).

We begin by defining what is meant by the probability generating function and the convolution of nonnegative integer-valued random variables.

<u>Definition 4.1.1</u>: Let $\{p_k\}$, $k = 0,1,2,\cdots$, be the probability distribution for a nonnegative integer-valued random variable X where $p_k = P[X = k]$. Then we say $f(s) = \sum_{k=o}^{\infty} p_k s^k$ is the probability generating function for X wherever the series converges.

Since $f(1) = 1$, $f(s)$ converges at least on $[-1,1]$.

<u>Definition 4.1.2</u>: Let X and Y be independent nonnegative integer-valued random variables with probability distributions $\{p_k\}$ and $\{q_k\}$ respectively. The sequence $\{h_k\}$ where $h_k = \sum_{i=1}^{k} p_i q_{k-i}$ is said to be the

convolution of $\{p_k\}$ and $\{q_k\}$ denoted $\{h_k\} = \{p_k\} * \{q_k\}$.

Theorem 4.1.1: If $W = X + Y$ where X, Y are as in Definition 4.1.2, $\{h_k\}$ is the probability distribution of W.

Proof: Since X, Y are independent we have $P[W = k] = P[X = 0]\cdot P[Y = k] + P[X = 1]P[Y = k - 1] + \cdots + P[X = k]P[Y = 0]$ or

$P[W = k] = \sum_{i=1}^{k} p_i q_{k-i} = h_k$. □

Theorem 4.1.2: If the probability generating functions associated with $\{p_k\}$ and $\{q_k\}$ are respectively $f(s)$ and $g(s)$, then the probability generating function associated with $\{h_k\} = \{p_k\} * \{q_k\}$ is $h(s) = f(s)g(s)$.

Proof: $f(s)g(s) = \left(\sum_{k=o}^{\infty} p_k s^k\right)\left(\sum_{k=o}^{\infty} q_k s^k\right)$ for $|s| \leq 1$.

Since both series converge absolutely, the Cauchy product converges to the product of the two series. Hence,

$$f(s)g(s) = \left(\sum_{k=o}^{\infty} p_k s^k\right)\left(\sum_{k=o}^{\infty} q_k s^k\right) = \sum_{k=o}^{\infty}\left(\sum_{i=o}^{k} p_i q_{k-i}\right)s^k = \sum_{k=o}^{\infty} h_k s^k = h(s).$$

□

The convolution of sequences is readily seen to extend to more than two sequences and we could have $\{h_k\} = \{p_k\} * \{q_k\} * \{r_k\}$, where the convolution operation is associative as well as commutative. Suppose we want the n-fold convolution of $\{p_k\}$ with itself. The notation often used is $\{h_k\} = \{p_k\} * \{p_k\} * \cdots * \{p_k\} = \{p_k\}^{n*}$.

A simple mathematical induction argument on Theorem 4.1.2 will show that if $h(s)$ is the probability generating function for $\{h_k\} = \{p_k\}^{n*}$, then $h(s) = [f(s)]^n$ where $f(s)$ is as defined in Theorem 4.1.2.

Theorem 4.1.3: Let $X_1, X_2, \cdots$ be a sequence of independent, identically distributed nonnegative integer-valued random variables with probability distribution $\{p_k\}$ and probability generating function $f(s)$. If N is a nonnegative integer-valued random variable having probability distribution $\{q_k\}$ and probability generating function $g(s)$, then the probability generating function of $S_N = \sum_{i=1}^{N} X_i$ is $h(s) = g(f(s))$, $0 < s \leq 1$.

Proof: Let $P[S_N = k] = h_k = \sum_{n=o}^{\infty} P[S_n = k]P[N = n]$.
$P[S_n = k] = P[X_1 + X_2 + \cdots + X_n = k] = \{p_k\}^{n*}$ by the extension of Theorem 4.1.1 using induction. Then we have $h_k = \sum_{n=o}^{\infty} q_n\{f_k\}^{n*}$ since $P[N = n] = q_n$. Hence, the probability generating function of S_N becomes $h(s) = \sum_{k=o}^{\infty} h_k s^k = \sum_{k=o}^{\infty}\left(\sum_{n=o}^{\infty} q_n\{p_k\}^{n*}\right)s^k = \sum_{n=o}^{\infty} q_n \sum_{k=o}^{\infty} \{p_k\}^{n*}s^k$ where the interchange is allowed by nonnegativity of all summands. By the extension of Theorem 4.1.2 previously alluded to, $\sum_{k=o}^{\infty} \{p_k\}^{n*}s^k = [f(s)]^n$. Then we have $h(s) = \sum_{n=o}^{\infty} q_n[f(s)]^n = g(f(s))$ since $|f(s)| \leq 1$. □

Corollary 4.1.1: If in Theorem 4.1.3, N is distributed identically to X_i, the probability generating function for S_N is $h(s) = f(f(s))$.

Proof: Simply replace g by f in Theorem 4.1.3 since the probability generating function of N is also $f(s)$. □

The previous corollary will be employed in the analysis of the probability theory application known as cascade or branching processes.

The Galton-Watson branching processes concern the regeneration properties of objects which in applications may be neutron particles, bacteria cells, family names, genes, customers in a waiting line or other physical objects. We shall use the term particles to represent these objects.

Let $Z_0, Z_1, Z_2, \cdots$ be a sequence of random variables where Z_n represents the number of particles in the n^{th} generation. In words, the basic assumptions of the process are:

(1) Beginning with a single particle, each particle is able to create like particles.

(2) Every particle has probability p_k of creating k new particles.

(3) The particles of each generation act independently of one another.

Mathematically, the sequence $\{Z_k\}$ is a Markov chain since the size of any generation depends only on the preceding generation size.

Let $Z_{n,m}$ be the number of offspring from particle m of the n^{th} generation. Hence, $Z_1 = Z_{o,1}$ since there is only one particle for the zeroth generation. Then $Z_2 = Z_{1,1} + Z_{1,2} + Z_{1,3} + \cdots + Z_{1,Z_1}$ and inductively

$$Z_n = \sum_{m=1}^{Z_{n-1}} Z_{n-1,m} \tag{4.1.1}$$

By Corollary 4.1.1, if F(s) is the probability generating function for the sequence $\{p_k\}$, then the distribution of Z_2 has probability generating function $F(F(s)) = F^{(2)}(s)$; $0 < s \leq 1$. Continuing this process and noting that the domain and range of $F(s)$ are the same, we obtain the result that Z_n has probability generating function $F^{(n)}(s)$, the n^{th} iterate of $F(s)$.

Basic assumptions are that $E[Z_1] = b$, $1 < b < +\infty$ and to insure a finite variance for Z_1, let $\sum_{k=0}^{\infty} k^2 p_k < +\infty$. To avoid trivial cases we shall assume that $p_k \neq 1$ for all k and $p_0 + p_1 < 1$.

The probability generating function for Z_n is thus given by $F^{(0)}(s) \equiv s$ and $F^{(n)}(s) = F(F^{(n-1)}(s))$ for $0 < s \leq 1$ and $n = 1,2,\cdots$.

Using Theorem 25 from Buck (1965), page 199, we may express the mean and variance of Z_1, and hence those of Z_n, in terms of first and second left-derivatives of F at $s = 1$. If Var Z_1 is denoted by σ^2 we may then obtain the mean and variance of Z_n as follows:

$$E[Z_n] = b^n \text{ and Var } Z_n = \frac{b^n(b^n - 1)}{b(b - 1)} \sigma^2.$$

Define the sequence of scaled random variables $\{W_n\}$ where $W_n = Z_n/b^n$. From the preceding paragraph we find that the mean and variance of W_n are

$$E[W_n] = 1 \text{ and Var } W_n = \frac{b^n(b^n - 1)}{b(b - 1)} \cdot \frac{\sigma^2}{b^{2n}} < K.$$

Note that Var $W_n < K$ implies $E[|W_n^2|] < L$ which by Doob (1953), page 629, is a sufficient condition for uniform integrability of $\{W_n\}$. (The reference cited also defines this concept.)

Harris (1963), page 14, indicated the possibility of the following argument.

Definition 4.1.3: If $\{Y_n\}$ is an arbitrary sequence of random variables where $E[|Y_n|] < +\infty$ and $E[Y_{n+1}|Y_n,Y_{n-1},\cdots,Y_1,Y_o] = Y_n$ almost surely for all n, $\{Y_n\}$ is a martingale.

Theorem 4.1.3: $\{W_n\} = \{Z_n/b^n\}$ is a martingale.

Proof: $Z_o,Z_1,\cdots$ constitute a Markov chain so $E[Z_{n+1}|Z_n,Z_{n-1},\cdots, Z_1,Z_o] = E[Z_{n+1}|Z_n]$. Since $E[Z_{n,m}] = b$, $E[Z_{n+1}|Z_n] = bZ_n$ almost surely. Then $E[W_{n+1}|W_n] = E[W_{n+1}|Z_n]$ since the σ-algebra generated by W_n is equivalent to that generated by Z_n. Hence, $E[W_{n+1}|W_n] = E[Z_{n+1}/b^{n+1}|Z_n]$ $= Z_n/b^n = W_n$ almost surely. Therefore, we conclude that $\{W_n\}$ is a martingale. □

By the martingale convergence theorem in Doob (1953), page 319, if $E[|W_n|]$ is uniformly bounded, then $W_n \to W$ almost surely. Furthermore, since $\{W_n\}$ is uniformly integrable we have also that $E[W_n] \to E[W]$, so $E[W] = 1$.

Let $G_n(w)$ be the distribution function of W_n. Since $W_n \geq 0$, the moment generating function $\emptyset_n(s) = \int_0^\infty e^{sw}dG_n$ of W_n exists on $(-\infty,0]$. Since W_n converges almost surely to W, it converges in distribution and if G(w) is the distribution function of W we have the moment generating function $\emptyset(s) = \int_0^\infty e^{sw}dG(w)$ of W existing on $(-\infty,0]$ since W must be nonnegative. Then by the Helly-Bray theorem $\emptyset_n(s) \overset{n}{\to} \emptyset(s)$ and $\emptyset'(0) = E[W] = 1$ in view of the analog for integrals, of Theorem 25, Buck (1965), page 199.

Using the properties of moment generating functions

$$\emptyset_n(sb) = F^{(n)}[e^{s/b^{n-1}}] = F[F^{(n-1)}(e^{s/b^{n-1}})];\ -\infty < s \leq 0. \quad (4.1.2)$$

Upon rewriting (4.1.2) we obtain

$$\emptyset_n(sb) = F[\emptyset_{n-1}(s)]; \ -\infty < s \le 0. \tag{4.1.3}$$

By continuity of F on (0,1] and taking limits of both sides as $n \to \infty$ gives

$$\emptyset(sb) = F(\emptyset(s)); \ -\infty < s \le 0. \tag{4.1.4}$$

Equation (4.1.4) is seen to be identical with equation (3.1.1) with $g(s) = bs$. This result is seen in Harris (1963), page 15.

$\emptyset(s)$ is a moment generating function that is continuous and strictly increasing on $(-\infty,0]$ so $\emptyset^-(t)$ exists on $0 < t \le 1$. Letting $t = \emptyset(s)$ in (4.1.4) gives

$$F(t) = \emptyset[b(\emptyset^-(t))]; \ 0 < t \le 1. \tag{4.1.5}$$

The n^{th} iterate of $F(t)$ may then be expressed by

$$F^{(n)}(t) = \emptyset[b^n\emptyset^-(t)]; \ 0 < t \le 1. \tag{4.1.6}$$

By centering at 1 and rescaling by b^{-n} we let $t = 1 + \frac{s}{b^n}$ in (4.1.6) which becomes

$$F^{(n)}(1 + \frac{s}{b^n}) = \emptyset[b^n\emptyset^-(1 + \frac{s}{b^n})]; \ -\infty < s \le 0. \tag{4.1.7}$$

A modification of Lemma 4.3.6 yields a left-derivative of $\emptyset^-$ at 1 so setting $H = \emptyset^-$, (4.1.7) gives

$$F^{(n)}(1 + \frac{s}{b^n}) = \emptyset[b^nH(1) + H'(1) \cdot s + b^nR(\frac{s}{b^n})]; \ -\infty < s \le 0$$

$$\text{where } \frac{R(\frac{s}{b^n})}{(\frac{s}{b^n})} \xrightarrow{n} 0 \text{ or } b^nR(\frac{s}{b^n}) \xrightarrow{n} 0 \text{ for each } s. \tag{4.1.8}$$

By continuity of $\emptyset$, we take limits in (4.1.8) and get

$$\lim_{n \to \infty} F^{(n)}(1 + \frac{s}{b^n}) = \emptyset[\frac{1}{\emptyset'(0)} \cdot s]. \tag{4.1.9}$$

However, $\emptyset(0) = 1$ so $\emptyset^{-}(1) = 0$ and $\emptyset'(0^{-}) = 1 = E[W]$. Hence, assuming $\emptyset'(0) = \emptyset'(0^{-})$, we have

$$\lim_{n \to \infty} F^{(n)}(1 + \frac{s}{b^n}) = \emptyset(s); \ -\infty < s \le 0. \tag{4.1.10}$$

By (4.1.5), F(t) was an easily iterated function with edge $\emptyset$ and core b where F(t) is the probability generating function of Z_1, $E[Z_1] = b > 1$ and $\emptyset(s)$ is the moment generating function of the limit random variable W. By (4.1.10), wherever it is defined, we see that $\emptyset$ is the limit of suitably scaled iterates of F. This is an additional relationship between the probability generating function and moment generating function which comes about through iteration and scaling, which although simple to obtain, does not seem to be a well known result.

To illustrate the idea we shall observe a simple numerical example using a single point distribution. This was one of the trivial cases excluded by our branching process assumptions, but since it is easy to see how the limit function is obtained, we shall include it.

Example 4.1.1: Let $Z_n = 2^n$, $P[Z_n = 2^n] = 1$. Then $E[Z_1] = 2^1$, $E[Z_2] = 2^2, \cdots, E[Z_n] = 2^n$. Let $W_n = 2^{-n}Z_n$ for $n = 1,2,3,\cdots$. Then $W_n \equiv 1$ so $\emptyset(s) = e^s$ and $F(t) = t^2$. Setting $t = 1 + 2^{-n}s$ gives $F^{(n)}(1 + 2^{-n}s) = (1 + 2^{-n}s)^{2^n} \overset{n}{\to} e^s = \emptyset(s)$.

In the next section we shall consider an extension of the preceding results with integer valued random variables, to a branching process analog having continuous random variables.

B. A Continuous Extension

Consider a family of real random variables X_ν with distributions G_ν parametrized by a real parameter ν. Suppose that the natural parameter range R includes unity and includes as well the supports S_ν of all distributions G_ν, $\nu \varepsilon R$ stated by $1 \varepsilon R$ and $\bigcup_{\nu \varepsilon R} S_\nu \subset R$. In other words, the set of ν-values for which G_ν is a distribution includes all values "possible" under all distributions and also unity.

Suppose in addition that each X_ν is absolutely continuous, with density $g_\nu(\cdot)$ with respect to some measure $\mu(\cdot)$ and possesses a Mellin transform

$$M_\nu(t) = \int t^x g_\nu(x)d\mu(x),\ 0 \leq t \leq 1$$

in which the parameter ν enters exponentially:

$$M_\nu(t) = [M_1(t)]^\nu \tag{4.2.1}$$

Examples of this can be found among the infinitely divisible distributions, for example, the Gamma family, for which of course $R = S_\nu = (0,+\infty)$.

Consider now the Markov chain $Y_1, Y_2, \cdots$ with densities $f_{Y_1,Y_2,\cdots,Y_n}(y_1,y_2,\cdots,y_n) = \prod_{i=1}^{n} g_{y_{i-1}}(y_i)$ with respect to $\prod_{i=1}^{n} \mu(dy_i)$, where $Y_0 \equiv 1$. This chain, whose structure is indicated by writing $Y_{i+1} = X_{Y_i}$, is analogous to an ordinary branching process. In particular, the Mellin transform $M^n(\cdot)$ of Y_n is the n^{th} iterate $M_1^{(n)}(\cdot)$ of the Mellin transform $M_1(\cdot)$ of G_1.

For example,

$$M^3(t) \equiv \int_{y_3} t^{y_3}[\int_{y_1}\int_{y_2} g_1(y_1)g_{y_1}(y_2)g_{y_2}(y_3)d\mu(y_1)d\mu(y_2)]d\mu(y_3)$$

$$= \int_{y_1}\int_{y_2} g_1(y_1)g_{y_1}(y_2)[\int_{y_3} t^{y_3} g_{y_2}(y_3)d\mu(y_3)]d\mu(y_1)d\mu(y_2)$$

$$= \int_{y_1}\int_{y_2} g_1(y_1)g_{y_1}(y_2)[M_{y_2}(t)]d\mu(y_1)d\mu(y_2). \qquad (4.2.2)$$

By (4.2.1) we may write (4.2.2) as

$$M^3(t) \equiv \int_{y_1}\int_{y_2} g_1(y_1)g_{y_1}(y_2)[M_1(t)]^{y_2}d\mu(y_2)d\mu(y_1). \qquad (4.2.3)$$

Then $M^3(t) \equiv \int_{y_1} g_1(y_1)\{\int_{y_2} g_{y_1}(y_2)[M_1(t)]^{y_2}d\mu(y_2)\}d\mu(y_1)$

$$= \int_{y_1} g_1(y_1)[M_{y_1}(M_1(t))]d\mu(y_1). \qquad (4.2.4)$$

Applying (4.2.1) to (4.2.4) we have

$$M^3(t) \equiv \int_{y_1} g_1(y_1)[M_1(M_1(t))]^{y_1}d\mu(y_1)$$

which by definition is

$$M^3(t) \equiv M_1(M_1(M_1(t))) \equiv M^{(3)}(t). \qquad (4.2.5)$$

If $G_1(\cdot)$ has an expectation $b > 1$ and a finite variance, the arguments and conclusions of the previous section apply.

C. A Theorem on Iteration

In preparation for the theorem we shall exhibit a sequence of definitions, facts and lemmas.

<u>Definition 4.3.1</u>: If $L(x)$ is a real-valued, strictly increasing function on a subset D of the reals, define the set D' by

$$D' = \{u : \inf_{x\varepsilon D} L(x) < u < \sup_{x\varepsilon D} L(x)\}.$$

<u>Definition 4.3.2</u>: If $L(x)$ is a strictly increasing function on a subset D of the reals, define the function $L^*(u)$ on D' by

$$L^*(u) = \sup_{x\varepsilon D}\{x : L(x) \leq u\}.$$

<u>Definition 4.3.3</u>: Let $L(a^-)$ and $L(a^+)$ denote the left and right-hand limits of $L(x)$ at a, respectively. $L(x)$ has a jump type discontinuity at a if $L(a^-)$ and $L(a^+)$ are both finite but $L(a^-) \neq L(a^+)$.

The following facts are well known and presented without proof. (See for example, Rudin (1953), page 72).

<u>Fact (A)</u>: The set of points at which a monotone function is discontinuous is at most countable.

<u>Fact (B)</u>: Any discontinuities of a monotone function must be of the jump type.

The following sequence of eight lemmas will refer to the functions L and L* as previously defined.

<u>Lemma 4.3.1</u>: L^* is a left inverse for L on any domain D for which L is strictly increasing.

<u>Proof</u>: By the definition of L^* we have $L^*[L(x)] = \sup_{y \varepsilon D}\{y : L(y) \leq L(x)\}$ and since L is strictly increasing on D, this implies $L^*[L(x)] = x$ for all $x \varepsilon D$. $\square$

Let I be the interval $I = \{x : -\delta < x < \delta\}$ where $\delta > 0$. Let I' be the interval $\{u : L(-\delta) < u < L(\delta)\}$. With these definitions of I and I' we state the next lemma.

<u>Lemma 4.3.2</u>: If L is strictly increasing and continuous on I, then L^* is a right inverse of L on I'.

<u>Proof</u>: By Lemma 4.3.1, $L^*[L(x)] = x$ for $x \varepsilon I$. Since L is strictly increasing on I we have that $L[L^*(L(x))] = L(x)$ for $x \varepsilon I$. Hence, $L[L^*(u)] = u$ for all $u \varepsilon L(I)$, i.e., by continuity, for all $u \varepsilon I'$. $\square$

<u>Lemma 4.3.3</u>: L^* is a nondecreasing function on D'.

<u>Proof</u>: Let $u,v \varepsilon D'$ such that $u < v$. Then $\{x : L(x) \leq u\}$ $\{x : L(x) \leq v\}$ so we have $\sup_{x \varepsilon D}\{x : L(x) \leq u\} \leq \sup_{x \varepsilon D}\{x : L(x) \leq v\}$. Hence, by the definition of L^*, $L^*(u) \leq L^*(v)$. $\square$

<u>Lemma 4.3.4</u>: If $L(x)$ has a jump discontinuity at $x = a$, i.e. $L(a^-) < L(a^+)$, then $L^*(u) = a$ for all $u \varepsilon E'$ where $E' = \{u : L(a^-) < u < L(a^+)\}$.

Proof: Suppose $L^*(u) \neq a$ for some $u \in E'$, say $u = v$. Then $L^*(v) = a + k$ for some $k \neq 0$. Choose any ε such that $0 < \varepsilon < |k|$. Then we have the inequalities

$$L(a - \varepsilon) \leq L(a^-) < v < L(a^+) \leq L(a + \varepsilon).$$

Since L^* is a left inverse for L and is nondecreasing by Lemma 4.3.3 we have after applying L^*

$$a - \varepsilon \leq L^*(v) \leq a + \varepsilon.$$

However, by hypothesis $L^*(v) = a + k$ where $\varepsilon < |k|$ so we have a contradiction. Therefore $L^*(u) = a$ for all $u \in E'$. $\square$

Lemma 4.3.5: If L is strictly increasing on $D = \{x : -\delta < x < \delta\}$, then L^* is continuous on $D' = \{u : \inf_{x \in D} L(x) < u < \sup_{x \in D} L(x)\}$.

Proof: Suppose L^* is discontinuous at some point of D', say v. Then by Fact (b) and Lemma 4.3.3 we have the following inequalities:

$L^*(v - \varepsilon) \leq L^*(v^-) < L^*(v^+) \leq L^*(v + \varepsilon)$ for all $\varepsilon > 0$ such that $v - \varepsilon$ and $v + \varepsilon$ are in D'. Let $A_\varepsilon = \{x : L(x) < v + \varepsilon\}$ and $B_\varepsilon = \{x : L(x) \leq v - \varepsilon\}$. Define $V_\varepsilon = A_\varepsilon - B_\varepsilon = \{x : L^*(v - \varepsilon) < x < L^*(v + \varepsilon)\}$. By hypothesis $U = \{x : L^*(v^-) < x < L^*(v^+)\}$ is a fixed nondegenerate interval with $U \subset V_\varepsilon$ for every $\varepsilon > 0$. Moreover, on V_ε, $v - \varepsilon < L(x) < v + \varepsilon$ which implies $L(x) = v$ on U. This is contradictory to the assumption that $L(x)$ is strictly increasing on D. Therefore $L^*(u)$ is continuous for all $u \in D'$. $\square$

Lemma 4.3.6: If $L(x)$ is strictly increasing on $D = \{x : -\delta < x < \delta\}$ such that $L'(0) \neq 0$ exists and $L(0) = p$, then $L^{*\prime}(p) = \frac{1}{L'(0)}$.

Proof: Since $L'(0) \neq 0$ exists, $L(x)$ is continuous at $x = 0$ and L^* is locally a right inverse also. Then $L(0) = p$ is equivalent to $L^*(p) = 0$. Strict monotonicity of L implies $x \to 0$ is equivalent to $L(x) \to L(0)$. Let $u = L(x)$ so $L^*(u) = x$ and we have

$$L^{*\prime}(p) = \lim_{u \to p} \frac{L^*(u) - L^*(p)}{u - p} = \lim_{L(x) \to L(0)} \frac{x - 0}{L(x) - L(0)}$$

$$= \lim_{x \to 0} \frac{1}{\frac{L(x) - L(0)}{x - 0}} = \frac{1}{L'(0)}. \quad \square$$

Lemma 4.3.7: Let $L(x)$ be strictly increasing on D such that $L(0) = p$. Let $b > 1$ be a real number and n a positive integer. Then $L'(0) = L^{*\prime}(p) = 1$. if and only if $b^n L^*(p + \frac{x}{b^n}) \overset{n}{\to} x$ uniformly for $x \in I^*$, where $I^* = \{x : 0 \leq |x| < \beta\}$ for some β such that $0 < \beta < 1$.

Proof: Let $L'(0) = L^{*\prime}(p) = 1$. Then given $\varepsilon > 0$ there exists a $\delta_\varepsilon > 0$ such that for $0 < |h| < \delta_\varepsilon$ we have

$$\left| \frac{L^*(p + h) - L^*(p)}{h} - 1 \right| = \left| \frac{L^*(p + h)}{h} - 1 \right| < \varepsilon.$$

Then there exists an integer N_ε such that $n > N_\varepsilon$ implies that $\beta / b^n < \delta_\varepsilon$. Hence for all $x \in I = \{x : 0 < |x| < \beta\}$ and $n > N_\varepsilon$

$$\sup_{x \in I} \left| \frac{L^*(p + \frac{x}{b^n})}{\frac{x}{b^n}} - 1 \right| = \sup_{x \in I} \frac{1}{|x|} \left| b^n L^*(p + \frac{x}{b^n}) - x \right| < \varepsilon.$$

Since $\frac{1}{|x|} > 1$ for $x \in I$ we conclude that $b^n L^*(p + \frac{x}{b^n}) \overset{n}{\to} x$ uniformly on I.

For the converse, define the set I' by $I' = \{x : 0 < \alpha \leq |x| \leq \beta < 1\}$ where α is such that $\alpha b < \beta$. Then we are given that

$\lim_{n \to \infty} \sup_{x \in I} \left| b^n L^*(p + \frac{x}{b^n}) - x \right| = 0$ which implies

$$\lim_{n \to \infty} \sup_{x \in I'} |b^n L^*(p + \frac{x}{b^n}) - x| = 0 \text{ since } I' \subset I^*. \tag{4.3.1}$$

Now $|b^n L^*(p + \frac{x}{b^n}) - x| = \left| \frac{L(p + \frac{x}{b^n})}{\frac{x}{b^n}} - 1 \right| |x| \geq \alpha \left| \frac{L^*(p + \frac{x}{b^n})}{\frac{x}{b^n}} - 1 \right|$

for all $x \in I'$ so we have

$$\sup_{x \in I'} |b^n L^*(p + \frac{x}{b^n}) - x| \geq \alpha \sup_{x \in I'} \left| \frac{L^*(p + \frac{x}{b^n})}{\frac{x}{b^n}} - 1 \right|.$$

Then by (4.3.1) $\lim_{n \to \infty} \sup_{x \in I'} \left| \frac{L^*(p + \frac{x}{b^n})}{\frac{x}{b^n}} - 1 \right| = 0.$ (4.3.2)

In view of (4.3.2), given $\varepsilon > 0$, there exists $M(\varepsilon)$ such that $n > M(\varepsilon)$ implies

$$\sup_{x \in I'} \left| \frac{L^*(p + \frac{x}{b^n})}{\frac{x}{b^n}} - 1 \right| < \varepsilon.$$

Now if $\delta(\varepsilon) \equiv \beta / b^{M(\varepsilon)}$, we claim that $0 < |y| < \delta(\varepsilon)$ implies that

$$\left| \frac{L^*(p + y)}{y} - 1 \right| < \varepsilon, \text{ i.e. } L^{*\prime}(p) = 1.$$

This because if there exists a y such that $0 < |y| < \delta(\varepsilon) \equiv \beta / b^{M(\varepsilon)}$, since $\alpha < \beta / b$, there is an $N \geq M(\varepsilon)$ such that $\frac{\alpha}{b^N} \leq |y| \leq \frac{\beta}{b^N}$. Then for

$J = \{y : \alpha/b^N \leq |y| \leq \beta/b^N\}$; we have

$$\sup_{y \in J}\left|\frac{L^*(p + y)}{y} - 1\right| = \sup_{x \in I'}\left|\frac{L^*(p + \frac{x}{b^N})}{\frac{x}{b^N}} - 1\right| < \varepsilon. \quad \square$$

Lemma 4.3.8: If $f(u) = L[bL^*(u)]$ is a strictly increasing function on $N' = \{u : p - \varepsilon < u < p + \varepsilon\}$ then $L(x)$ is continuous on $L^*(N')$, a neighborhood of zero.

Proof: Suppose that L is discontinuous at some point $d \in L^*(N')$. By Fact (B) the discontinuity is of the jump type and if $E' = \{u : L(d^-) < u < L(d^+)\}$, E' is nondegenerate. By Lemma 4.3.4, $L^*(u) = d$ on $E' \cap N'$ so $bL^*(u) = bd$ is constant for $u \in E' \cap N'$. Then $f(u) = L[bL^*(u)] = L(bd)$ is constant on $E' \cap N'$. Since $E' \cap N'$ is nondegenerate, the strict monotonicity property of $f(u)$ is contradicted and $L(x)$ must be continuous at all points of $L^*(N')$. $\square$

We now state the theorem of this section.

Theorem 4.3.1: Let $L(x)$ be strictly increasing on N, a neighborhood of zero and let $L(0) = p$. Define $f(u) = L[bL^*(u)]$.

For a number $b > 1$,

(i) $f^{(n)}(p + \frac{x}{b^n}) \overset{n}{\to} L(x)$ uniformly on a deleted neighborhood of zero;

(ii) $f(u) = L[bL^*(u)]$ is strictly increasing on a neighborhood of p;

if and only if $L(x)$ is continuous on a neighborhood of zero and $L'(0) = 1$.

Proof: Let $N = \{x : -\varepsilon < x < \varepsilon\}$ for $\varepsilon > 0$ be the neighborhood on which $L(x)$ is strictly increasing. Then for $L^*(u)$ as in Definition 4.3.2, we have L^* continuous and nondecreasing on the neighborhood $N' = \{u : L(-\varepsilon) < u < L(\varepsilon)\}$. For $0 < \delta < \varepsilon$, define the interval $N_o = \{x : -\delta < x < \delta\}$. Let $N_k' = \{u : L(-\varepsilon/b^k) < u < L(\varepsilon/b^k)\}$ be neighborhoods of u defined for $k = 1,2,3,\cdots$.

Given that $L(x)$ is continuous on N_o and $L'(0) = 1$, by Lemma 4.3.6 $L^{*\prime}(p) = 1$ and by Lemma 4.3.7, $b^n L^*(p + \frac{x}{b^n}) \overset{n}{\to} x$ uniformly for $x \in N$, assuming without loss of generality, that $\varepsilon < 1$. Then on N_o, a subset of N, the convergence is uniform and for $x \in N_o$, $b^n L^*(p + \frac{x}{b^n})$ is eventually in N_o when n is sufficiently large. Since L is continuous on N_o we have upon taking the limit,

$$\lim_{n \to \infty} L[b^n L^*(p + \frac{x}{b^n})] = L[\lim_{n\to\infty} b^n L^*(p + \frac{x}{b^n})] = L(x). \qquad (4.3.3)$$

with the convergence uniform on N_o. Now we see $f(u) = L[bL^*(u)]$, $u \in N_1'$. Clearly $f(u)$ is easily iterated and we may write

$$f^{(n)}(u) = L[b^n L^*(u)]; \; u \in N_n'. \qquad (4.3.4)$$

Replacing u by $p + \frac{x}{b^n}$ in (4.3.4) yields

$$f^{(n)}(p + \frac{x}{b^n}) = L[b^n L^*(p + \frac{x}{b^n})]. \qquad (4.3.5)$$

Letting $x \in N_o$ and applying (4.3.3) we obtain

$$f^{(n)}(p + \frac{x}{b^n}) \overset{n}{\to} L(x) \text{ uniformly on } N_o. \qquad (4.3.6)$$

It remains to show that $f(u)$ is strictly increasing on a neighborhood of p. For this, observe the set $N_1' = \{u : L(-\frac{\delta}{b}) < u < L(\frac{\delta}{b})\}$. Since L is strictly increasing on N_o and continuous also, by Lemma 4.3.2 L^* is an inverse on $L(N_o)$. By Lemma 4.3.6, $L'(x) > 0$ for $x \varepsilon N_o$ is equivalent to $L^{*\prime}(u) > 0$ for $u \varepsilon L(N_o)$. $N_1' \subset L(N_o)$ and for $u \varepsilon N_1'$ we obtain $x = bL^*(u)$ is in N_o. Then $bL^*(u) = bL^-(u)$ is a strictly increasing function so $f(u) = L[bL^*(u)]$ is a strictly increasing function of a strictly increasing function and hence is itself strictly increasing on N_1'.

For the converse, define $M' = \{u : p - \delta < u < p + \delta\}$ as the set on which $f(u) = L[bL^*(u)]$ is strictly increasing for some $b > 1$. By Lemma 4.3.8, $L(x)$ is continuous on $M = L^*(M')$, so L^* is an inverse of L for $x \varepsilon M$. Let $M_o = \{x : 0 < |x| \leq \beta\}$ where $\beta < \varepsilon < 1$ be the neighborhood on which $f^{(n)}(p + \frac{x}{b^n}) \overset{n}{\to} L(x)$ uniformly. Using the easily iterated property we obtain

$$f^{(n)}(u) = L[b^n L^*(u)]; \; u \varepsilon N_n'. \tag{4.3.7}$$

Setting $u = p + \frac{x}{b^n}$ in (4.3.7) yields

$$f^{(n)}(p + \frac{x}{b^n}) = L[b^n L^*(p + \frac{x}{b^n})]. \tag{4.3.8}$$

By hypothesis, $f^{(n)}(p + \frac{x}{b^n}) \overset{n}{\to} L(x)$ uniformly on M_o so (4.3.8) becomes

$$\lim_{n \to \infty} L[b^n L^*(p + \frac{x}{b^n})] = L(x) \text{ uniformly on } M_o. \tag{4.3.9}$$

L is continuous on M and is therefore continuous on $M_1 = M \cap M_o$, so (4.3.9) may be written

$$L[\lim_{n\to\infty} b^n L^*(p + \frac{x}{b^n})] = L(x) \text{ uniformly on } M_1. \tag{4.3.10}$$

Since L^* is a left inverse on N', applying L^* to (4.3.10) gives the desired results,

$$\lim_{n \to \infty} \sup_{x \varepsilon M_1} \left| b^n L^*(p + \frac{x}{b^n}) - x \right| = 0. \tag{4.3.11}$$

M_1 is a certain neighborhood of zero as required in Lemma 4.3.7 so we have $L'(0) = 1$. □

A slight modification in the theorem is possible by the following corollary.

<u>Corollary 4.3.1</u>: Statement (ii) of Theorem 4.3.1 may be replaced by (ii)' $f(u)$ is strictly increasing on a neighborhood of p and satisfies the functional equation $f^{(k)}[L(\frac{x}{b^k})] = L(x)$, $k = 1,2,3,\cdots$; $-\infty < x < +\infty$.

<u>Proof</u>: Let $N_\alpha = \{x : -\alpha < x < \alpha\}$ where $0 < \alpha < \delta$ so N_α is a subset of N_0, the continuity set for $L(x)$. Define $f(u)$ to be strictly increasing on the set $M_\alpha = \{u : L(-\frac{\alpha}{b}) < u < L(\frac{\alpha}{b})\}$. Suppose that f satisfies the functional equation of the corollary. Then it must satisfy it for the case $k = 1$ and where $x \in N_\alpha$. In this case we obtain

$$f[L(\tfrac{x}{b})] = L(x), \ x \in N_\alpha. \tag{4.3.12}$$

Since $N_\alpha \subset N_0$, L^* is an inverse, so by setting $x = bL^*(u)$ we have (4.3.12) becoming

$$f(u) = L[bL^*(u)], \ u \in M_\alpha. \tag{4.3.13}$$

Conversely, if $f(u) = L[bL^*(u)]$ is strictly increasing on M_α, by the easily iterated property (4.3.13) becomes for $L(-\alpha/b^k) < u < L(\alpha/b^k)$,

$$f^{(k)}(u) = L[b^k L^*(u)]; \ k = 1,2,3,\cdots. \tag{4.3.14}$$

Setting $u = L(\frac{x}{b^k})$ and since L^* is an inverse on M_α, $b^k L^*(u) = x$. This gives (4.3.14) as

$$f^{(k)}[L(\frac{x}{b^k})] = L(x);\ k = 1,2,3,\cdots;\ -\infty < x < +\infty. \qquad (4.3.15)$$

This because $x = b^k L^*(u)$ for $u \varepsilon M_\alpha$ implies $-b^{k-1}\alpha < x < b^{k-1}\alpha$ for $k = 1,2,3,\cdots$ or $-\infty < x < +\infty$. □

The functional equation (4.3.15) was discussed in Thomas and David (1968), with respect to proving continuity and strict monotonicity of the limit function. This corollary relates it to the easily iterated concept as we have defined it.

A further modification is possible in the next corollary.

Corollary 4.3.2: Statement (ii) of Theorem 4.3.1 may be replaced by (ii)'' $f(u)$ is strictly increasing on a neighborhood of p and f is continuous everywhere.

Proof: Using the now assumed continuity of f together with statement (i) of the theorem we may write

$$\lim_{n\to\infty} f^{(n)}(p + \frac{x}{b^n}) = f[\lim_{n\to\infty} f^{(n-1)}(p + \frac{x/b}{b^{n-1}})] = f[L(\frac{x}{b})] = L(x). \qquad (4.3.16)$$

Now (4.3.16) is the functional equation (4.3.15) with $k = 1$ and the result follows from Corollary 4.3.1. □

The next two corollaries give further characterization to the roles played by the numbers p and b in the theorem.

Corollary 4.3.3: The number p is the coordinate of a fixed point for f, i.e. $f(p) = p$.

Proof: $f(p) = L[bL^*(p)] = L[b \cdot 0] = L(0) = p$. □

<u>Corollary 4.3.4</u>: $f(u)$ is differentiable at p and $f'(p) = b$.

<u>Proof</u>: $L'(0) = 1$ and $L^{*\prime}(p) = 1$ imply that $f'(p) = L'[bL^*(p)] \cdot b \cdot L^{*\prime}(p) = L'(0) \cdot b \cdot L^{*\prime}(p) = b$. □

<u>Corollary 4.3.5</u>: $L(x) = p + x + o(x)$ for x in some neighborhood of 0.

<u>Proof</u>: $L(x) - L(0) = xL'(0) + o(x)$ where $L(0) = p$ and $L'(0) = 1$. □

To illustrate the use of the preceding theorem and corollaries we shall show how it relates to the maximin function, $\emptyset(x)$, which will be derived in the next section. Thomas and David (1968), established that $\emptyset^{(n)}(a + \frac{x}{b^n})$ eventually approached $L(x)$ monotonically and that $L(x)$ was continuous and strictly increasing everywhere. $\emptyset(t)$ had the interval $[0,1]$ as its domain and range and was continuous there. Dini's theorem states that if $\{g_n\} \overset{n}{\to} g$ monotonically where $g_1, g_2, \cdots, g$ are continuous on a closed bounded interval, the convergence is uniform on that interval. Since the function $\emptyset(t)$ and all its iterates are continuous on $[0,1]$ as is $L(x)$, by Dini's theorem the convergence is uniform on $[0,1]$. Since $0 < a < 1$ where $\emptyset(a) = a$, then clearly $\emptyset^{(n)}(a + \frac{x}{b^n})$ converges uniformly to $L(x)$ on a determined neighborhood of zero. Applying Corollary 4.3.2 gives us the result that $L'(0) = 1$.

Furthermore, $\emptyset(t) = L[bL^*(t)]$, i.e. $\emptyset(t)$ is an easily iterated function. This was implied by the fact that it was stated to play the role of the f function in the functional equation (4.3.14). However, it was never written in this particular form and the concept of "easily iterated" was not developed. Many characteristics of $L(x)$ were obtained with regard to convexity, monotonicity and continuity in the reference cited.

D. Further Considerations and Examples of Convergence of Scaled Iterates

In this section we shall exploit some of the classical results from iteration of analytic functions of a complex variable to determine characteristics of the limit function $L(x)$ in a real variable setting.

The functional equation known as Schroeder's equation is

$$M[f(z)] = \beta M(z). \tag{4.4.1}$$

M and f are functions of a complex variable and β is a complex constant. The fundamental result proved by Koenigs (1884), is given by the following theorem.

Theorem 4.4.1: If $g(z)$ is analytic at 0 with $g'(0) = \beta$, $0 < |\beta| < 1$, then $\lim_{n \to \infty} \beta^{-n} g^{(n)}(z) = M(z)$ where $M(z)$ is analytic at 0 and satisfies $M[g(z)] = \beta M(z)$ with $M'(0) = 1$, provided $g(0) = 0$.

At this point we recall the translation procedure from Chapter III. If $f(z)$ is analytic at $z = \alpha$ where $f(\alpha) = \alpha$, we may set $h(z) = z - \alpha$ so $g(z) = h[f(h^{-}(z))]$ is analytic at 0 and $g(0) = 0$. Also, since $g(z) = f(z + \alpha) - \alpha$, $f'(\alpha) = \beta$ implies $g'(0) = \beta$ and we may examine $f(z)$ in view of Theorem 4.4.1 by using $g(z)$.

Theorem 4.4.2: Let $\phi(t)$ be a strictly increasing and continuous function of a real variable t on the interval [0,1] having fixed points at 0, a, 1 where $0 < a < 1$ and $\phi'(a) = b > 1$. Let $\phi^{-}(t) = f(t)$ where $g(z) = f(z + a) - a$ satisfies Theorem 4.4.1 when t is replaced by a complex variable z. Then for $x \in R_1$, $\lim_{n \to \infty} \phi^{(n)}(a + \frac{x}{b^n}) = L(x)$ where $L'(0) = 1$ and $L(x)$ is analytic at 0.

Proof: By Theorem 4.4.1, $\lim_{n \to \infty} b^n g^{(n)}(t) = M(t)$, since $g'(0) = \frac{1}{b}$, where $M'(0) = 1$ and $M(t)$ is analytic at 0. $M(g(0)) = \frac{1}{b}M(0)$ implies $M(0) = 0$ and $M'(0) = 1$ implies M^{-} exists in a neighborhood of 0.

Convergence everywhere of a sequence of real functions implies convergence in distribution of a corresponding sequence of random variables. Let $X = t + a$ where X is uniformly distributed on [0,1]. Then the limit statement $b^n g^{(n)}(t) \to M(t)$ may be expressed probabilistically by the convergence in distribution statement where $x \in R_1$,

$$P[b^n g^{(n)}(X - a) \leq x] \xrightarrow{D} P[M(X - a) \leq x]. \tag{4.4.2}$$

For X in a suitable neighborhood of a and x in a suitable neighborhood of 0 we have M^{-} existing and (4.4.2) becomes upon noting $g(t) = f(t + a) - a$,

$$P[b^n(f^{(n)}(X) - a) \leq x] \xrightarrow{D} P[X - a \leq M^{-}(x)]. \tag{4.4.3}$$

Recalling that $\phi^{-}(t) = f(t)$, we have (4.4.3) becoming

$$P[X \leq \phi^{(n)}(a + \frac{x}{b^n})] \xrightarrow{D} P[X \leq M^{-}(x) + a]. \tag{4.4.4}$$

However, $\emptyset^{(n)}(a + \frac{x}{b^n})$ is on [0,1] and since X is uniformly distributed on [0,1] we obtain

$$\lim_{n \to \infty} \emptyset^{(n)}(a + \frac{x}{b^n}) = M^-(x) + a = L(x). \tag{4.4.5}$$

M(x) is analytic at 0 and $M'(0) = 1$ so $M^-(x)$ is analytic at 0 implying L(x) is analytic at 0 with $L'(0) = 1$. □

<u>Corollary 4.4.1</u>: If $\emptyset(t)$ satisfies Theorem 4.4.2, then $\emptyset(t) = L[bL^-(t)]$ on some neighborhood of a.

<u>Proof</u>: Since $M(g(t)) = \frac{1}{b}M(t)$ where g(t) is $\emptyset^-(t + a) - a$, as in Theorem 4.4.2, the existence of M^- on a neighborhood of 0 gives

$$g(t) = M^-[b^{-1}M(t)]. \tag{4.4.6}$$

Since $g(t) = h[f(h^-(t))]$ for $h(t) = t - a$, we have the following result upon taking h^-,

$$f(h^-(t)) = h^- \circ M^-[b^{-1}M(t)]. \tag{4.4.7}$$

Taking the inverse of both sides yields

$$h(\emptyset(t)) = M^-[b(M \circ h)(t)]$$

or

$$\emptyset(t) = (M \circ h)^-[b(M \circ h)(t)]. \tag{4.4.8}$$

Letting $L^-(t) = (M \circ h)(t)$ in (4.4.8) gives the result, since $L(t) = h^-(M^-(t)) = M^-(t) + a$ as previously defined in (4.4.5). □

It should be noted that in application of the preceding theorem and corollary in forthcoming examples, the real variable theorem from the previous section could be applied. It does, however, require more work to establish the uniform convergence using Dini's theorem, etc. as illustrated by the prior discussion of the maximin function. The arguments leading to monotonicity of convergence to $L(x)$ are given in Thomas and David (1968), page 246.

Example 4.4.1: (The iterated median).

Suppose $X_1, X_2, X_3, \cdots, X_{3^n}$ constitutes a random sample of random variables distributed uniformly on the interval $[0,1]$. Note there are 3^n random variables in the sequence. If we take the median of consecutive groups of three we obtain the following sequence:

$$X_1^{(1)} = \text{med}(X_1, X_2, X_3),\ X_2^{(1)} = \text{med}(X_4, X_5, X_6), \cdots,\ X_{3^{n-1}}^{(1)} = \text{med}(X_{3^n-2}, X_{3^n-1}, X_{3^n}).$$

Continue the process on this sequence and obtain:

$$X_1^{(2)} = \text{med}(X_1^{(1)}, X_2^{(1)}, X_3^{(1)}),\ X_2^{(2)} = \text{med}(X_4^{(1)}, X_5^{(1)}, X_6^{(1)}), \cdots,$$
$$X_{3^{n-2}}^{(2)} = \text{med}(X_{3^{n-1}-2}^{(1)}, X_{3^{n-1}-1}^{(1)}, X_{3^{n-1}}^{(1)}).$$

If we iteratively continue this procedure we arrive at $X_1^{(n)} = \text{med}(X_1^{(n-1)}, X_2^{(n-1)}, X_3^{(n-1)})$. It is the distribution function of $X_1^{(n)}$ we desire.

Let $F_1(t) = P[X_i^{(1)} \le t] = P[X_1, X_2, X_3 \le t] + 3P[X_1, X_2 \le t]P[X_3 > t]$ or $F_1(t) = t^3 + 3t^2(1 - t) = -2t^3 + 3t^2$.

Then $F_2(t) = P[X_i^{(2)} \le t] = [F_1(t)]^3 + 3[F_1(t)]^2[1 - F_1(t)]$, so $F_2(t) = F_1[F_1(t)] = F_1^{(2)}(t)$.

Likewise, $F_n(t) = F_1(F_1^{(n-1)}(t)) = F_1^{(n)}(t)$ so we have the distribution function of the iterated median as the iterated distribution function of the median at the first stage.

Observing $F_1(t) = -2t^3 + 3t^2$ we see that it is continuous and strictly increasing on [0,1] with fixed points at 0, $\frac{1}{2}$, 1. $F_1'(\frac{1}{2}) = \frac{3}{2}$ and applying Theorem 4.3.2 with $\emptyset(t) = F_1(t)$ we have

$\lim_{n \to \infty} F_1^{(n)}(\frac{1}{2} + (\frac{2}{3})^n x) = L(x)$ where $L(x)$ is analytic at 0 and $L'(0) = 1$.

Example 4.4.2: (The iterated maximin).

Let $X_1, X_2, X_3, \cdots, X_{4^n}$ be a sequence of independent random variables distributed uniformly on [0,1]. First, we take the minimums of consecutive pairs and obtain the sequence $X_1^1 = \min(X_1, X_2), X_2^1 = \min(X_3, X_4), \cdots, X^1_{2^{2n-1}} = \min(X_{4^n-1}, X_{4^n})$. Then from the resulting sequence we take maximums of consecutive pairs to get $X_1^{(1)} = \max(X_1^1, X_2^1), X_2^{(1)} = \max(X_3^1, X_4^1), \cdots, X^{(1)}_{4^{n-1}} = \max(X^1_{2^{2n-1}-1}, X^1_{2^{2n-1}})$. If we continue this two stage procedure, we ultimately arrive at $X_1^{(n)}$ whose distribution function we wish. Let X, Y be independent uniform random variables on [0,1]. $H(t) = P[\min X,Y \leq t] = 1 - (1 - t)^2$ while $G(t) = P[\max X,Y \leq t] = t^2$ so $F_1(t) = G(H(t)) = [1-(1-t)^2]^2$ which is the maximin function previously referred to as $\emptyset(x)$. If we repeat the procedure, we have $F_2(t) = G(H(F_1(t))) = F_1^{(2)}(t)$ and finally $F_n(t) = F_1^{(n)}(t) = P[X_1^{(n)} \leq t]$ is the distribution function of the n^{th} iterate of the maximin which is the n^{th} iterate of $F_1(t)$.

If we took the minimums on groups of p random variables and the maximums on groups of q random variables at the second stage, the distribution function for the maximin is $\emptyset(t) = [1 - (1 - t)^p]^q$. This general case is discussed in Thcmas (1965), with respect to bound functions. The function $F_1(t) = [1 - (1 - t)^2]^2$ will be called the maximin function (with $p = q = 2$) although it is actually the distribution function of the maximin random variable $X_1^{(1)}$ on $[0,1]$. It is easily seen to be continuous and strictly increasing on $[0,1]$.

Fixed points may be determined to be 0, $\frac{3 - \sqrt{5}}{2}$, 1 on $[0,1]$. Let $a = \frac{3 - \sqrt{5}}{2}$ and it may be seen that $F_1'(a) = b > 1$ so F_1^- is analytic at a and we may apply Theorem 4.4.2. In this case, we obtain that $L(x)$ possesses all derivatives at 0 (in addition to the fact that $L'(0) = 1$) which extends the result of Theorem 4.3.1 with respect to analyticity of $L(x)$ in this situation.

Example 4.4.3: (The iterated minimax).

Simply reversing the roles of taking maximums and minimums in the previous example is called the minimax function and would be obtained by finding $W_1(t) = H(G(t)) = 1 - (1 - t^2)^2$. Similarly, the n^{th} iterate of the minimax has distribution function $W_n(t) = W^{(n)}(t)$. The symmetry about the median is seen by observing the fixed points at 0, $\frac{\sqrt{5} - 1}{2}$, 1 on $[0,1]$. The interior fixed point is complementary to that of the maximin since they add to 1. A short discussion of this and related ideas is given in Appendix A.

Again, since $W_1(t)$ is strictly increasing and continuous on $[0,1]$ and for $a = \frac{\sqrt{5} - 1}{2}$, $W_1'(a) = b > 1$, we have the results of Theorem 4.4.2 and Corollary 4.4.1 available.

Example 4.4.4: (The iterated maximed).

We have established the procedure of starting with a sequence of independent uniform random variables on [0,1]. If we first took the medians of groups of three and then the maximums of groups of two we can construct the distribution function $V_1(t) = 9t^4 - 12t^5 + 4t^6$. $V'(0) = V'(1) = 0$ so since there is a unique inflection point at $t = \frac{1}{9}$ we conclude there is a unique fixed point on (0,1). Clearly, $V_1(0) = 0$, $V_1(1) = 1$ and $V_1(t)$ is strictly increasing and continuous on [0,1] so the usual argument applies.

Since these examples are combinations of order statistics and hence are themselves order statistics, we are quite naturally led to consider the following example.

Example 4.4.4: (The incomplete beta function).

$$\text{Let } \emptyset(t) = \frac{1}{B(m,n)} \int_0^t x^{m-1}(1 - x)^{n-1}dx;\ 0 \leq t \leq 1;\ m > 1,\ n > 1.$$

$B(m,n)$ is the beta integral, so $\emptyset(t)$ is the incomplete beta distribution function. We shall show that $\emptyset(t)$ satisfies the requirements of Theorem 4.4.2 and Corollary 4.4.1.

First, we observe that $\emptyset(0) = 0$, $\emptyset(1) = 1$ and $\emptyset(t)$ is continuous and strictly increasing on [0,1]. Since $\emptyset'(t) = \frac{1}{B(m,n)} t^{m-1}(1 - t)^{n-1}$, it is clear that $\emptyset'(0) = \emptyset'(1) = 0$. Then for some $\varepsilon > 0, 0 \leq t < \varepsilon$ implies $\frac{\emptyset(t) - \emptyset(0)}{t - 0} = \emptyset'(\xi) < 1$ for $0 \leq \xi < t$ by the mean value theorem and continuity of $\emptyset'(t)$. Hence, $\emptyset(t) < t$ on $[0,\xi)$. A similar argument gives $\emptyset(t) > t$ on $(\delta,1]$ for some $\delta > 0$. Since $\emptyset(t)$ is strictly increasing there is at least one fixed point on (0,1). If there is a unique inflection point on (0,1), then the fixed point is unique.

$$\phi''(t) = \frac{1}{B(m,n)} t^{m-2}(1 - t)^{n-2}[t(2 - n - m) - (1 - m)].$$

Then $\phi''(t) = 0$ at $t = \frac{m - 1}{m + n - 2}$ is the unique inflection point on (0,1) since $m,n > 1$. Let the unique fixed point be $t = a$. Since $\phi(t) > t$ for $t > a$ and $\phi(t) < t$ for $t < a$, $\phi'(a) > 1$. Since $\phi(z)$ is analytic at $z = a$ for $t = z$, a complex variable, we have the result.

This leads us to the next example.

Example 4.4.5: (The k^{th} order statistic).

Let $X_1, X_2, \cdots, X_{m^n}$ be a sequence of independent random variables from the uniform distribution on [0,1]. Let Y_k be the k^{th} order statistic from a group of m of the X_i where $k < m$. The distribution function is $F(t) = P[Y_k \leq t] = \sum_{i=k}^{m} \binom{m}{i} t^i (1 - t)^{m-i}$. This may be expressed in terms of the incomplete beta function by $F(t) = \frac{1}{B(k, m - k + 1)} \int_0^t x^{k-1}(1 - x)^{m-k}dx$.

By the procedures employed in the previous examples, if we take the n^{th} iterate of the k^{th} order statistic, we obtain as its distribution function the n^{th} iterate of $F(t)$. Since $F(t)$ has the properties necessary to apply Theorem 4.4.2, as exhibited in Example 4.4.5, we have results similar to the iterated median.

In the following chapter, for notational convenience, we shall use the symbol $\varphi(t)$ rather than $\phi(t)$ as used in this section.

V. FURTHER TOPICS IN UNIVARIATE ITERATION

A. Bound Functions and Almost Sure Convergence

In previous probabilistic examples we found that distribution functions of iteratively composited functions of certain random variables were themselves iterative compositions of an original distribution function, say $\phi(t)$. It was ascertained that $\phi(t)$ was easily iterated, that is, of form $\phi(t) = L[bL^{-}(t)]$ in a neighborhood of a fixed point. However, explicit expressions for the edge functions are usually not known so an alternate approach is to obtain bound functions. A reasonable conjecture is that the bound functions should have a common fixed point with ϕ at $t = a$, $0 < a < 1$, and that they should have the same derivative at a as ϕ, the function to be bounded. For iterative investigation, the bound functions should also be easily iterated with known edge functions. This approach was employed by Thomas and David (1968), to characterize the limit function near the fixed point of the maximin function.

A further use of bound functions is to establish almost sure convergence in certain cases. We shall illustrate this idea by first finding bound functions for the distribution function of the iterated median from Example 4.4.1.

Recall that $F_1(t) = -2t^3 + 3t^2$; $0 \leq t \leq 1$. Let us denote $F_1(t)$ by $\phi(t)$. The fixed point on (0,1) is $a = 1/2$ and $b = \phi'(\frac{1}{2}) = \frac{3}{2}$.

If we use the exponential type of bound functions as exhibited in Example 3.2.6 we have $\lambda(t) = 1 - \frac{1}{2}[2(1 - t]^{3/2}$ and $\mu(t) = \frac{1}{2}(2t)^{3/2}$. These were seen to be easily iterated giving

$$\lambda^{(n)}(t) = 1 - \frac{1}{2}[2(1 - t)]^{(3/2)^n} \text{ and } \mu^{(n)}(t) = \frac{1}{2}(2t)^{(3/2)^n}. \quad (5.1.1)$$

We shall first show that $\lambda(t) \leq \phi(t) \leq \mu(t)$ for $0 \leq t \leq 1$. Starting with verification of $\lambda(t) \leq \phi(t)$ we have

$$1 - \frac{1}{2}[2(1 - t)]^{3/2} \leq -2t^3 + 3t^2. \tag{5.1.2}$$

Simplifying (5.1.2) gives

$$[2(1 - t)]^{3/2} \geq 2(2t^3 - 3t^2 + 1) \tag{5.1.3}$$

The derivative of the right hand side is negative for $t < 1$ so it is a decreasing function that is 0 at $t = 1$, therefore it is nonnegative on $[0,1]$. We may then square both sides of (5.1.3).

$$2(1 - t)^3 \geq (2t^3 - 3t^2 + 1)^2 \tag{5.1.4}$$

Expanding both sides yields, after transposing terms

$$D_\lambda(t) = -4t^6 + 12t^5 - 9t^4 - 6t^3 + 12t^2 - 6t + 1 \geq 0.$$

That $D_\lambda(t) \geq 0$ is established by noting that the zeros of $D_\lambda(t)$ occur at $t = -1, \frac{1}{2}$ or 1 so on $[0,1]$ the only zeros are $\frac{1}{2}$ and 1. Now $D_\lambda(0) = \sqrt{2} > 0$ and $D_\lambda(\frac{3}{4}) = \frac{4\sqrt{2} - 5}{32} > 0$ so since the only sign change could occur at $t = 1/2$ we are assured that $D_\lambda(t) \geq 0$ on $[0,1]$. Reversing the steps gives the validity of (5.1.2).

Applying the same approach to $\phi(t) \leq \mu(t)$ we start with

$$-2t^3 + 3t^2 \leq \frac{1}{2}(2t)^{3/2}. \tag{5.1.5}$$

Squaring both sides gives, after algebraic simplification

$$t(4t^2 - 12t + 9) \leq 2$$

or $$D_\mu(t) = 4t^3 - 12t^2 + 9t - 2 \leq 0. \tag{5.1.6}$$

All zeros of $D_\mu(t)$ are at 1/2 and 2 so any sign change on [0,1] must occur at $t = 1/2$.

$D_\mu(\frac{1}{4}) = -\frac{7}{16} < 0$ and $D_\mu(1) = -2 < 0$ so clearly (5.1.5) holds on [0,1] and we have the result

$$\lambda(t) \leq \phi(t) \leq \mu(t) \text{ on } 0 \leq t \leq 1. \tag{5.1.7}$$

Noting that λ, ϕ and μ are strictly increasing on [0,1] we have

$$\lambda(\lambda(t)) \leq \lambda(\phi(t)) \leq \phi(\phi(t)) \leq \mu(\phi(t)) \leq \mu(\mu(t))$$

or $$\lambda^{(2)}(t) \leq \phi^{(2)}(t) \leq \mu^{(2)}(t) \text{ on } [0,1].$$

Proceeding in this way, inductively we see that

$$\lambda^{(n)}(t) \leq \phi^{(n)}(t) \leq \mu^{(n)}(t) \text{ on } [0,1]. \tag{5.1.8}$$

If we denoted the random variable Z_n to be the n^{th} iterated median ($X_1^{(n)}$ in Example 4.4.1), we may show that Z_n converges almost surely to 1/2.

Theorem 5.1.1: If Z_n is a sequence of random variables such that $P[Z_n \leq t] = \phi^{(n)}(t)$ where $\phi(t) = -2t^3 + 3t^2$, then $Z_n \to \frac{1}{2}$ almost surely.

Proof: Let ε be given such that $0 < \varepsilon < \frac{1}{2}$. We then have the inequalities as follows:

$$P[Z_n \leq \tfrac{1}{2} - \varepsilon] = \phi^{(n)}(\tfrac{1}{2} - \varepsilon) \leq \mu^{(n)}(\tfrac{1}{2} - \varepsilon) = \tfrac{1}{2}(1 - 2\varepsilon)^{(3/2)^n} \tag{5.1.9}$$

$$P[Z_n > \tfrac{1}{2} + \varepsilon] = 1 - \phi^{(n)}(\tfrac{1}{2} + \varepsilon) \leq 1 - \lambda^{(n)}(\tfrac{1}{2} + \varepsilon) = \tfrac{1}{2}(1 - 2\varepsilon)^{(3/2)^n}. \tag{5.1.10}$$

Combining (5.1.9) and (5.1.10) gives

$$P[|Z_n - \tfrac{1}{2}| > \varepsilon] \leq (1 - 2\varepsilon)^{(3/2)^n}. \tag{5.1.11}$$

Since $\varepsilon < \frac{1}{2}$, summing both sides yields

$$\sum_{n=1}^{\infty} P[|Z_n - \tfrac{1}{2}| > \varepsilon] \leq \sum_{n=1}^{\infty} (1 - 2\varepsilon)^{(3/2)^n} \leq \sum_{n=1}^{\infty} (1 - 2\varepsilon)^n < +\infty. \quad (5.1.12)$$

Now employing the Borel-Cantelli lemma we find that by virtue of (5.1.12)

$$P[|Z_n - \tfrac{1}{2}| > \varepsilon \text{ infinitely often}] = 0.$$

This is equivalent to saying that there exists an integer $N > 0$ such that

$$P[|Z_n - \tfrac{1}{2}| \leq \varepsilon \text{ for all } n > N] = 1. \quad (5.1.13)$$

Then, $P[\lim_{n\to\infty} Z_n = \frac{1}{2}] = 1$. $\square$

The previous discussion shows that if we can use easily iterated bound functions that lead to convergent series as in Equation (5.1.12), we can employ this technique. It must first be established that the functions bound ϕ over the whole interval $[0,1]$ which is usually difficult to do.

It was shown by Thomas (1965), that if Y_n is the maximin random variable with distribution function $\phi(t) = [1 - (1 - t)^2]^2$, then the functions $\lambda(t) = 1 - (1 - a)\left(\frac{1 - t}{1 - a}\right)^b$ and $\mu(t) = a\left(\frac{t}{a}\right)^b$ will bound $\phi(t)$ on $[0,1]$. As before, $\phi'(a) = b$ where a is the unique fixed point on $(0,1)$.

Using the steps of the previous argument we have the next theorem.

<u>Theorem 5.1.2</u>: If Y_n is a sequence of random variables such that $P[Y_n \leq t] = \phi^{(n)}(t)$ where $\phi(t) = [1 - (1 - t)^2]^2$, then $Y_n \to a$ almost surely with $a = \frac{3 - \sqrt{5}}{2}$.

By the symmetry about $t = 1/2$ of the maximin and minimax functions, we may also include the minimax random variable. Clearly the same bound functions will serve by using the appropriate fixed point, $a' = \frac{\sqrt{5} - 1}{2}$ and we have the following theorem.

Theorem 5.1.3: If W_n is a sequence of random variables such that $P[W_n \leq t] = \phi^{(n)}(t)$ where $\phi(t) = 1 - (1 - t^2)^2$, then $W_n \to a'$ almost surely, with $a' = \frac{\sqrt{5} - 1}{2}$.

We are led to conjecture as to the limitations of the use of λ and μ in proving almost sure convergence of iterated order statistic random variables. Taking Z_n to be the 3^{rd} order statistic from consecutive groups of size four in a sample sequence of independent random variables $X_1, X_2, \cdots, X_{4^n}$, uniform on [0,1], we obtain $\psi(t) = -3t^4 + 4t^3$ as the distribution function of Z_1. However, it may be seen, after some computation, that λ and μ do not bound ψ on [0,1]. Noting that the fixed point for ψ on (0,1) is a = .75 whereas the fixed point for the maximin was approximately .65, we speculate as to a "cutoff point" between .65 and .75 (and symetrically to .50, between .25 and .35). Another unresolved problem is whether or not all order statistic random variables whose distribution function has a fixed point between $\frac{\sqrt{5} - 1}{2}$ and $\frac{3 - \sqrt{5}}{2}$ can be treated by using the λ and μ bound functions in the preceding manner.

Another example of strong law type of convergence obtained by iterative considerations is to simply define a random variable as a convergent easily iterated function of a given random variable. For example, if $Z_1 = f(X) = \frac{2X}{X + 1}$, $Z_2 = f^{(2)}(X), \cdots, Z_n = f^{(n)}(X)$ we have $Z_n = \frac{2^n X}{(2^n - 1)X + 1}$. Obviously, $Z_n \to 1$ regardless of the distribution of X.

B. Nonstationary Aspects

Suppose that $F_k(t) = L[b_k L^*(t)]$, $k = 1,2,3,\cdots$ where L and L^* do not depend on b_k. Let $b_k > 1$ for all k and impose the condition that the infinite product, $\prod_{k=1}^{\infty} b_k$, diverges. From Titchmarsh (1939), page 14, a necessary and sufficient condition for this is to require that $\sum_{k=1}^{\infty} (b_k - 1)$ is a divergent series.

Let $G_n(t)$ be the n^{th} composition taken in the following order:

$$G_n(t) = F_n(F_{n-1}(\cdots F_2(F_1(t))\cdots)) = L[b_1 b_2 \cdots b_k L^*(t)].$$

Denote $\prod_{k=1}^{n} b_k$ by b_{n*} and assume p is a fixed point for each $F_k(t)$ such that $F_k'(p) = b_k$. Letting b_{n*} play the role of b^n in the earlier scaling procedure let $t = p + \frac{x}{b_{n*}}$. Then assuming $L'(0) = 1$ we have

$$G_n(p + \frac{x}{b_{n*}}) = L[b_{n*} L^*(p + \frac{x}{b_{n*}})]. \tag{5.2.1}$$

Using $L^*(p + \frac{x}{b_{n*}}) = L^*(p) + L^{*\prime}(p)\frac{x}{b_{n*}} + o(\frac{x}{b_{n*}})$ where $L^*(p) = 0$ and $L^{*\prime}(p) = L'(0) = 1$, (5.2.1) becomes, after taking limits as before

$$\lim_{n \to \infty} G_n(p + \frac{x}{b_{n*}}) = L[x + \lim_{n\to\infty} b_{n*}\, o(\frac{x}{b_{n*}})] = L(x). \tag{5.2.2}$$

Under very special conditions we may then apply the analogous limit procedures to composition of functions in the nonstationary situation as in the stationary situation of iteration, previously discussed.

Example 5.2.1: Referring to the bound functions of Example (3.2.6), suppose we have

$$\lambda_k(t) = 1 - (1 - a)\left(\frac{1 - t}{1 - a}\right)^{b_k};\ 0 \le t \le 1;\ b_k > 1,\ k = 1,2,3,\cdots$$

and

$$\mu_k(t) = a\left(\frac{t}{a}\right)^{b_k};\ 0 \le t \le 1;\ b_k > 1;\ k = 1,2,3,\cdots.$$

The n^{th} composition of these functions become respectively

$$\lambda_n \circ \lambda_{n-1} \circ \cdots \circ \lambda_2 \circ \lambda_1(t) = 1 - (1 - a)\left(\frac{1 - t}{1 - a}\right)^{b_{n^*}}$$

and

$$\mu_n \circ \mu_{n-1} \circ \cdots \circ \mu_2 \circ \mu_1(t) = a\left(\frac{t}{a}\right)^{b_{n^*}}.$$

Note that since the linear fractional functions have edge functions depending on b_k, we are unable to use this procedure so it is limited to easily iterated functions of identical form.

VI. APPROXIMATE AND CONDITIONAL INVERSES

A. Definitions and Properties

To obtain certain results with iteration of mappings in R_n it will be necessary to employ the concept of matrix rescaling. Since we do not wish to unduly restrict our discussion to nonsingular core matrices we shall appeal to the idea of approximate and conditional inverses of linear transformations.

Suppose we are given the quadruple (M,A,V,U) where:

(1) V, U are subspaces of R_n such that

$V \cap U = \{0\}$ and $V \oplus U = R_n$

and (2) M, A are linear transformations of R_n into itself.

In view of (1) every vector $x \in R_n$ has a unique representation, $x = u + v$, where $u \in U$ and $v \in V$. Unique, because $x = u + v$ and $x = u' + v'$ imply $u - u' = v - v'$, hence if $u \neq u'$ and $v \neq v'$ then $u - u' = v - v' = a \neq 0$. But this contradicts $a \in V \cap U$ which contains only the null vector.

Definition 6.1.1: Given a matrix M, if A is any matrix such that $MAM = M$, then A is a conditional inverse of M.

Definition 6.1.2: If for all vectors $x \in R_n$, there exists a vector $v_x \in R_n$, $v_x \neq 0$, such that $\lim_{k \to \infty} A^k M^k(x) = v_x$ then A is an approximate inverse of M.

(We shall not distinquish between left or right approximate inverses, since we could as well say if $\lim_{k \to \infty} M^k B^k(x) = v_x$ then B is an approximate inverse and proceed in the following discussion with minor modifications in the proofs and examples.)

Definition 6.1.3: Let N_M be the null space of M and let V be a subspace of R_n such that $V \cap N_M = \{0\}$. If $v \in V$ implies that $M(v) \in V$, then V will be said to be a nonsingular invariant subspace with respect to M which we shall denote nisM.

Suppose that M_V is the restriction of M to V, that is, $M_V(v) \equiv M(v)$ for $v \in V$. Then for V, a nisM, M_V is a nonsingular transformation of V onto itself. This may be seen as follows: Suppose $M_V(x) = 0$ for some $x \in V$. Then since $0 \in V$, $M_V(x) = M(x) = 0$. This implies $x \in N_M$ so $x \in V \cap N_M$ and hence $x = 0$. Therefore M_V is nonsingular since if it maps any vector of V into the zero vector, that vector must itself be the zero vector.

Since M_V is nonsingular, it has an inverse M_V^{-1} and we make the following definition.

Definition 6.1.4: Any A such that $A_V = M_V^{-1}$ is said to be an inverse of M relative to V.

Theorem 6.1.1: If A is an inverse of M relative to R_M, the range space of M, then

(i) A is an approximate inverse of M

and (ii) A is a conditional inverse of M.

Proof: We first prove (ii) by observing that for $x = u + v$, $AM(x) = AM(u + v) = AM(v) = A_V M_V(v) = v$ where $V = R_M$. Hence, $M(x) = M(v)$ for all $x \varepsilon \mathcal{R}_n$. Then it follows that $MAM(x) = M(v) = M(x)$ for all $x \varepsilon \mathcal{R}_n$.

The proof of (i) follows by induction since $A^k M^k(x) = v$ for $x \varepsilon \mathcal{R}_n$ implies that $A^k M^k$ is the identity transformation on $V = R_M$. The proof of (ii) led to the fact that AM is the identity transformation on R_M so we will show that $A^{k+1}M^{k+1}$ is also the identity on R_M assuming that $A^k M^k(x) = v$.

$A^{k+1}M^{k+1}(x) = A^{k+1}M^{k+1}(v) = A(A^k M^k)M(v) = AM(v)$ since $M(v) \varepsilon R_M$. Therefore $A^{k+1}M^{k+1}(x) = v$ for all $x \varepsilon \mathcal{R}_n$. □

B. Examples of Construction of Approximate Inverses

Example 6.2.1: Let M be an $m \times m$ matrix having m distinct characteristic roots λ_i such that $\lambda_i \neq 0$ for $i = 1,2,\cdots,m$ and $|\lambda_1| > |\lambda_2| > \cdots > |\lambda_m|$. Let the characteristic vectors corresponding to these roots and spanning $\mathcal{R}_m$ be the column vectors of the matrix $P = [P_1,P_2,\cdots,P_m]$. Denote the diagonal matrix having characteristic roots as elements by $\Lambda = [\mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_m)]$. If $P^{-1} = Q$ we have $M = P\Lambda Q$ whose ijth element is $\sum_{h=1}^{m} \lambda_h p_{ih} q_{hj}$. Then $M^k = P\Lambda^k Q = [m_{ij}(k)]$ where $m_{ij}(k) = \sum_{h=1}^{m} \lambda_h^k p_{ih} q_{hj}$. Let $A = P\Lambda^{-1}Q = M^{-1}$ so A is the exact inverse of M and $A^k M^k(x) = x$. This, then is a special case of an approximate inverse with $N_M = \{0\}$.

Example 6.2.2: Suppose in Example 6.2.1 not all $\lambda_i \neq 0$, that is, M is singular. In this case we may take $A = P[\lambda_1^{-1} I_m]Q$ where λ_1 is called the dominant root since $|\lambda_1| = \max_i |\lambda_i|$. Then $A^k M^k = P[\lambda_1^{-k} I_m]P^{-1} P \Lambda^k P^{-1}$ $= \lambda_1^{-k} P \Lambda^k P^{-1} = \lambda_1^{-k} M^k$. In the limit we have

$$\lim_{k\to\infty} \lambda_1^{-k} M^k = \lim_{k \to \infty} \left[\sum_{h=1}^{m} \left(\frac{\lambda_h}{\lambda_1}\right)^k p_{ih} q_{hj} \right] = \lim_{k \to \infty} \left[\sum_{h=2}^{m} \left(\frac{\lambda_h}{\lambda_1}\right)^k p_{ih} q_{hj} \right] + [p_{i1} q_{1j}]$$

$$= [p_{i1} q_{1j}] = [c_{ij}] = C.$$

Hence, $\lim_{k \to \infty} A^k M^k(x) = C(x) = v_x$. If $P_1, P_2, \cdots, P_m$ are column vectors of P and $Q_1, Q_2, \cdots, Q_m$ are row vectors of Q we have $v_x = P_1 Q_1(x)$. If we consider x as represented by the basis $P_1, P_2, \cdots, P_m$, then we may write

$v_x = P_1 Q_1 \left(\sum_{j=1}^{m} x_j P_j \right) = P_1 x_1$ since $Q_i P_j = \delta_{ij}$ (the Kronecker delta function). Obviously $v_x \in V$, the space V being spanned by P_1 in this case.

This example illustrates the fact that scalar norming of a matrix by the dominant root, a frequently used technique in matrix iterative methods, is simply a special case of rescaling by an approximate inverse.

Example 6.2.3: To illustrate situations intermediate to the extremes of the first two examples, suppose M has distinct characteristic roots and without loss of generality we consider M a 4 × 4 matrix. Let $|\lambda_1| > |\lambda_2| > |\lambda_3| > |\lambda_4|$ where corresponding characteristic vectors are P_1, P_2, P_3, P_4 spanning R_4. Let V be the nisM spanned by P_1, P_2 and let A be a matrix with simple roots $\lambda_1^{-1}, \lambda_2^{-1}, \delta_3, \delta_4$ where $\max(|\delta_3|, |\delta_4|) < \min(|\lambda_3^{-1}|, |\lambda_4^{-1}|)$. Let corresponding characteristic vectors be P_1, P_2, D_3, D_4. If P_3, P_4 span U, then $U \cap V = \{0\}$ and $U \oplus V = R_4$. Let x be any vector in R_4. Then $M^k(x) = M^k(u) + M^k(v)$ where

$$M^k(v) = M^k(x_1P_1 + x_2P_2) = x_1\lambda_1^kP_1 + x_2\lambda_2^kP_2. \quad (6.1.1)$$

Multiplying by A^k gives that for $k = 1,2,3,\cdots$ we have

$$A^kM^k(v) = A^k(x_1\lambda_1^kP_1 + x_2\lambda_2^kP_2) = x_1P_1 + x_2P_2. \quad (6.1.2)$$

Now consider the vector $M^k(u)$ which may be written

$$M^k(u) = M^k(x_3P_3 + x_4P_4) = x_3\lambda_3^kP_3 + x_4\lambda_4^kP_4. \quad (6.1.3)$$

Since P_3 and P_4 are linear combinations of P_1,P_2,D_3,D_4 we may replace P_3 by $a_1P_1 + a_2P_2 + a_3D_3 + a_4D_4$ and P_4 by $b_1P_1 + b_2P_2 + b_3D_3 + b_4D_4$ in (6.1.3) and multiply by A^k to obtain

$$\begin{aligned} A^kM^k(u) &= A^k[x_3\lambda_3^k(a_1P_1 + a_2P_2 + a_3D_3 + a_4D_4) \\ &\quad + x_4\lambda_4^k(b_1P_1 + b_2P_2 + b_3D_3 + b_4D_4)] \\ &= x_3[\lambda_1^{-k}\lambda_3^ka_1P_1 + \lambda_2^{-k}\lambda_3^ka_2P_2 + \delta_3^k\lambda_3^ka_3D_3 + \delta_4^k\lambda_3^ka_4D_4] \\ &\quad + x_4[\lambda_1^{-k}\lambda_4^kb_1P_1 + \lambda_2^{-k}\lambda_4^kb_2P_2 + \delta_3^k\lambda_4^kb_3D_3 + \delta_4^k\lambda_4^kb_4D_4] \quad (6.1.4) \end{aligned}$$

From our specification of δ_3 and δ_4 it is seen that (λ_3/λ_1), (λ_3/λ_2), $(\delta_3\lambda_3)$, $(\delta_4\lambda_3)$, (λ_4/λ_1), (λ_4/λ_2), $(\delta_3\lambda_4)$ and $(\delta_4\lambda_4)$ are all less than unity. Therefore, taking the limit as $k \to \infty$ in (6.1.4) we simply have $A^kM^k(u) \xrightarrow{k} 0$. Hence, for any $x \in R_4$, $A^kM^k(x) \xrightarrow{k} x_1P_1 + x_2P_2 = v_x$ and A is an approximate inverse of M relative to V.

It should be noted that the choice for D_3 and D_4 is immaterial in this construction as long as P_1,P_2,D_3,D_4 span R_4. Then we are free to use the original characteristic vectors P_3 and P_4 as D_3 and D_4 respectively.

M need not be nonsingular, in which case the characteristic roots of A corresponding to the zero roots of M are arbitrary. A simple numerical example illustrates this comment, referring to Example 6.2.3.

Suppose $\lambda_1 = 4$, $\lambda_2 = 3$, $\lambda_3 = 2$ and $\lambda_4 = 0$. Then we may write $M = P[\text{diag}(4,3,2,0)]P^{-1}$ and choose $A = P[\text{diag}(\frac{1}{4},\frac{1}{3},\frac{1}{2},\delta_4)]P^{-1}$. In this case $V = R_M$ the range space of M and A is also a conditional inverse by Theorem 6.1.1. Clearly, δ_4 is arbitrary and may be taken as zero. If we want V spanned by P_1,P_2 only, we may choose $A = P[\text{diag}(\frac{1}{4},\frac{1}{3},0,0)]P^{-1}$. In this case, A is an approximate inverse but not a conditional inverse.

It is clear that A may be chosen so that $M^kA^k(x) \to v_x$ and by an earlier comment, left or right approximate inverses may be considered in the preceding examples. Obviously, A is not unique if it is to satisfy only the requirement that it be an approximate inverse.

VII. ASYMPTOTIC ASPECTS OF MULTIVARIATE SCALING AND ITERATION

A. Scaling with the Differential Matrix in the Nonsingular Case

We are interested in obtaining results analogous to some of those in Chapter IV extended from R_1 to R_n. In R_1, the scaling constant was the derivative evaluated at a fixed point. The natural extension is the $n \times n$ differential matrix evaluated at a fixed point of the transformation. Hopefully, this matrix enjoys some of the scaling properties of its R_1 counterpart, however, as suggested by Karlin and McGregor (1970), there are certain difficulties in this approach. Restrictions needed for sufficient conditions become a bit cumbersome, but we shall exhibit some approaches to the problem.

We shall first state two definitions and a theorem concerning matrices. These may be found in Varga (1962), where the proof of the theorem is given on page 13.

Definition 7.1.1: An $m \times m$ matrix A is said to be convergent if the sequence $\{A^n\}$ converges to the null matrix.

Definition 7.1.2: If the $m \times m$ matrix A has characteristic roots $\lambda_1, \lambda_2, \cdots, \lambda_m$ we will call $\max_i |\lambda_i|$ the spectral radius of A, denoted $\rho(A)$.

Theorem 7.1.1: If B is an $m \times m$ matrix, B is convergent if and only if $\rho(B) < 1$.

We shall now establish useful criteria for the analysis to be used in this chapter.

Let $T(x)$ be an easily iterated mapping from D onto itself, D a subset of R_m, where $T(x) = \emptyset A \emptyset^*(x)$ has the following properties:

(1) $T(p) = p$.

(2) $T'(p) = A$ where A^{-1} exists and $\min_i |\lambda_i| > 1$, λ_i the characteristic roots of A.

(3) $\emptyset^*$ is of class $C_1(E)$ and $\emptyset$ is of class $C_1(F)$, where E and F are open sets in R_m containing the points p and 0, respectively.

If we denote $B = A^{-1}$, by (2) we have $\rho(B) < 1$ so $B^n \to 0$ (the null matrix) from Theorem 7.1.1. Since $\lambda_i \neq 1$ for all i, we see that $|A - I| \neq 0$ so $A - I$ is a nonsingular matrix. Hence, by Theorem 3.1.3, $\emptyset^*(p) = 0$.

By nonsingularity of A, $|A| = |T'(p)| \neq 0$ so from the chain rule we see that

$$|\emptyset'(A\emptyset^*(p))||A||\emptyset^{*\prime}(p)| = |\emptyset'(0)||A||\emptyset^{*\prime}(p)| \neq 0.$$

Then $\emptyset^{*\prime}(p)$ and $\emptyset'(0)$ are nonsingular so $\emptyset^*$ is a local inverse of $\emptyset$ on a neighborhood of p. Furthermore, $\emptyset(0) = p$ and we have the matrix equality, $[\emptyset^{*\prime}(p)] = [\emptyset'(0)]^{-1}$.

By assumption, $T(x)$ is e.i. on D so we have

$$T^{(n)}(x) = \emptyset A^n \emptyset^*(x), \quad x \in D. \tag{7.1.1}$$

Centering at p and rescaling by B^n is equivalent to setting $x = p + B^n y$ in (7.1.1) which becomes

$$T^{(n)}(p + B^n y) = \emptyset A^n \emptyset^*(p + B^n y), \quad y \in R_m. \tag{7.1.2}$$

Since $\emptyset^*$ is of class $C_1(E)$ and $\emptyset^*(p) = 0$, for sufficiently large n we have $\emptyset^*(p + B^n y) = [\emptyset^{*\prime}(p)](B^n y) + R(B^n y)$. Then (7.1.2) becomes

$$T^{(n)}(p + B^n y) = \emptyset\{A^n[\emptyset^{*\prime}(p)](B^n y) + A^n R(B^n y)\}. \qquad (7.1.3)$$

Sufficient conditions for arriving at a limit map analogous to that in the R_1 case are that both of the following hold:

(i) Either A^n or B^n commutes with $\emptyset^{*\prime}(p)$.

(ii) $A^n R(B^n y) \overset{n}{\to} 0$ for all $y \in R_m$.

When (i) holds the right-hand side of (7.1.3) becomes

$$\emptyset\{[\emptyset^{*\prime}(p)]y + A^n R(B^n y)\}. \qquad (7.1.4)$$

Since $\emptyset$ is continuous on F at least, let G be the continuity set of $\emptyset$. Then if $[\emptyset^{*\prime}(p)]y$ is in the interior of G, when (ii) holds, for n sufficiently large we have that $[\emptyset^{*\prime}(p)]y + A^n R(B^n y)$ is in G. Therefore, in view of (7.1.4), we have upon taking limits in (7.1.3) the result,

$$\lim_{n \to \infty} T^{(n)}(p + B^n y) = \emptyset\{[\emptyset^{*\prime}(p)]y\}. \qquad (7.1.5)$$

We now prove a theorem concerning matrices to obtain sufficient conditions for (ii) to hold.

<u>Theorem 7.1.2</u>: Let A be an $m \times m$ nonsingular real matrix, $m \geq 2$, and $\{\lambda_i\}$ as the characteristic roots. Let $B = A^{-1}$ where we denote the ijth elements of A^n and B^n by $a_{ij}(n)$ and $b_{ij}(n)$ respectively. If the characteristic roots of A are such that if $1 < \min_i|\lambda_i| \leq \max_i|\lambda_i| < [\min_i|\lambda_i|]^2$, then $a_{ij}(n) \cdot b_{kl}(n) \cdot b_{rs}(n) \overset{n}{\to} 0$ for all possible values of the subscripts.

Proof: Let J be the Jordan cononical form for A. Then there exist matrices P and P^{-1} such that $A = P^{-1}JP$ or $J = PAP^{-1}$. Clearly, $A^n = p^{-1}J^nP$ and therefore $B^n = P^{-1}J^{-n}P$. In the following discussion "bounded above in absolute value" will be abbreviated to "bounded above". The elements of J are bounded above by the elements of J_*, a matrix having $|\lambda_1|,|\lambda_2|,\cdots,|\lambda_m|$ on the diagonal and having 1's on the subdiagonal.

Without loss of generality, let $|\lambda_1| = \max_i|\lambda_i|$ and $|\lambda_2| = \min_i|\lambda_i|$. After some manipulation it may be seen that the elements of J^n are bounded above by the corresponding elements of the matrix M(n) given by

$$M(n) = \begin{bmatrix} |\lambda_1^n| & \binom{n}{1}|\lambda_1^{n-1}| & \binom{n}{2}|\lambda_1^{n-2}| & \cdots & \binom{n}{m-1}|\lambda_1^{n-m+1}| \\ & |\lambda_1^n| & \binom{n}{1}|\lambda_1^{n-1}| & \cdots & \binom{n}{m-2}|\lambda_1^{n-m+2}| \\ & & |\lambda_1^n| & \cdots & \binom{n}{m-3}|\lambda_1^{n-m+3}| \\ & \text{(SYMMETRIC)} & & \ddots & \vdots \\ & & & & |\lambda_1^n| \end{bmatrix}.$$

When $n > |\lambda_1| + m - 1$, every element of M(n) is bounded above by $n^{m-1}|\lambda_1^{n-m+1}|$. Since we are interested in $A^n = P^{-1}J^nP$, we observe that premultiplying by P^{-1} and postmultiplying by P will not change the order of magnitude with respect to n. That is, each element of A^n is bounded above by $c_1n^{m-1}|\lambda_1^{n-m+1}|$ for some constant c_1.

By a similar argument we observe that $J^{-1} = Q^{-1}HQ$ where H is the Jordan form for J^{-1}. Elements of H are bounded above by elements of H*, a matrix having $|\lambda_1^{-1}|, |\lambda_2^{-1}|, \cdots, |\lambda_m^{-1}|$ on the diagonal and 1's on the subdiagonal. Elements of H^n are then bounded above by $n^{m-1}|\lambda_2^{-n+m-1}|$ so $B^n = (QP)^{-1}H^n(QP)$ has elements bounded above by $c_2 n^{m-1}|\lambda_2^{-n+m-1}|$ for some constant c_2. Then for all possible subscript combinations,

$$|a_{ij}(n) \cdot b_{kl}(n) \cdot b_{rs}(n)| \leq \frac{c_1 c_2^2 n^{3m-3}|\lambda_1^{n-m+1}|}{|\lambda_2^{n-m+1}|^2} \tag{7.1.6}$$

Denoting the right side of (7.1.6) by $U(n)$ we have the following order of magnitude relation

$$U(n) \sim n^d \left(\frac{|\lambda_1|}{|\lambda_2|^2}\right)^n. \tag{7.1.7}$$

Since d is a constant and by hypothesis $|\lambda_1| < |\lambda_2|^2$, $U(n) \overset{n}{\to} 0$ which gives the desired result. $\square$

Corollary 7.1.1: For the case $m = 2$ in Theorem 7.1.2, the condition that $1 < \min_i|\lambda_i| \leq \max_i|\lambda_i| < [\min_i|\lambda_i|]^2$ is a minimally nontrivial sufficient condition that $a_{ij}(n) \cdot b_{kl}(n) \cdot b_{rs}(n) \overset{n}{\to} 0$ for all possible values of the subscripts.

For the meaning of "minimally nontrivial sufficient conditions" as well as the type of details for the proof of the corollary the reader is referred to Appendix B. It should be noted that other sufficient conditions for the conclusion of Theorem 7.1.2 to hold are possible. These, however, involve relationships between all the characteristic roots rather than $\min_i|\lambda_i|$ and $\max_i|\lambda_i|$. We shall not pursue this approach further since we only wish to indicate a type of criteria under which condition (ii) would hold.

It will be demonstrated in Theorem 7.1.4 that for T(x) satisfying (1), (2) and (3) that condition (ii) will hold provided A satisfies the hypotheses of Theorem 7.1.2 (or Corollary 7.1.1 if m = 2). Condition (i) would not necessarily hold.

We shall first show that for a particular type of easily iterated map, Theorem 7.1.2 gives stronger results than needed for condition (ii) and furthermore that condition (i) is easily seen to hold.

<u>Theorem 7.1.3</u>: Let T(x) satisfy (1), (2) and (3) where $m \geq 2$ and T(x) is of the form

$$T(x) = \begin{bmatrix} \emptyset_1 & & \bigcirc \\ & \emptyset_2 & \\ \bigcirc & & \ddots \ \ \emptyset_m \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & & \vdots \\ \vdots & & & \\ a_{m1} & & \cdots & a_{mm} \end{bmatrix} \begin{bmatrix} \emptyset_1^* & & \bigcirc \\ & \emptyset_2^* & \\ \bigcirc & & \ddots \ \ \emptyset_m^* \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix}.$$

Let $\emptyset^*$ be of class $C_2(E)$ and let $\emptyset^{*\prime}(p) = cI$ for some scalar constant c. If $a_{ij}(n) \cdot b_{jl}(n) \cdot b_{lk}(n)$ for all possible i,j,k,l values, then

$$\lim_{n\to\infty} T^{(n)}(p + B^n y) = \emptyset(cy).$$

<u>Proof</u>: Since $\emptyset^*(p) = 0$ and $\emptyset^{*\prime}(p) = cI$ we have $\emptyset_i^*(p_i) = 0$ and $\emptyset_i^{*\prime}(p_i) = c$ for $i = 1,2,\cdots,m$. Condition (i) is clearly satisfied because A^n and cI commute. It remains to show condition (ii) holds. Let $x = p + B^n y$ in $T^{(n)}(x)$ gives $T^{(n)}(p + B^n y)$ whose i^{th} component we denote $t_i(n)$. Let the i^{th} row of A^n be $A_i^{(n)}$ and the j^{th} row of B^n be $B_j^{(n)}$.

Then we have

$$t_i(n) = \emptyset_i[A_i^{(n)}\emptyset^*(p + B^n y)]. \tag{7.1.8}$$

Writing (7.1.8) in summation notation yields

$$t_i(n) = \emptyset_i[\sum_{j=1}^{m} a_{ij}(n)\emptyset_j^*(p_j + B_j^{(n)}y)]. \tag{7.1.9}$$

Applying Taylor's approximation to (7.1.9) gives the term in the bracket in (7.1.9) as

$$\sum_{j=1}^{m} a_{ij}(n)[\emptyset_j^*(p_j) + \emptyset_j^{*\prime}(p_j)B_j^{(n)}y + \frac{\emptyset_i^{*\prime\prime}(\xi)}{2}(B_j^{(n)}y)^2]. \tag{7.1.10}$$

In (7.1.10), $p_j < \xi_j < p_j + B_j^{(n)}y$. By using $\emptyset_j^*(p_j) = 0$ and $\emptyset_j^{*\prime}(p_j) = c$ we have from (7.1.10),

$$c\sum_{j=1}^{m} a_{ij}(n)B_j^{(n)}y + \sum_{j=1}^{m} a_{ij}(n)(B_j^{(n)}y)^2\frac{\emptyset_j^{*\prime\prime}(\xi)}{2}. \tag{7.1.11}$$

The left sum in (7.1.11) becomes

$$\sum a_{ij}(n)B_j^{(n)}y = c\sum_{j=1}^{m} a_{ij}(n)\sum_{k=1}^{m} b_{jk}(n)y_k$$

$$= c\sum_{k=1}^{m}\sum_{j=1}^{m} a_{ij}(n)b_{jk}(n)y_k = c\sum_{k=1}^{m}\delta_{ik}y_k = cy_i.$$

(δ_{ik} is the Kronecker delta function.) (7.1.12)

The right sum of (7.1.11) will become,

$$\sum_{j=1}^{m} a_{ij}(n)(B_j^{(n)}y)^2\frac{\emptyset_j^{*\prime\prime}(\xi)}{2} = \sum_{j=1}^{m} a_{ij}(n)[\sum_{k=1}^{m} b_{jk}(n)y_k]^2\frac{\emptyset_j^{*\prime\prime}(\xi)}{2}$$

$$= \sum_{j=1}^{m}\sum_{k=1}^{m}\sum_{l=1}^{m}[a_{ij}(n)b_{jl}(n)b_{lk}(n)]y_{jl}y_{lk}\frac{\emptyset_j^{*\prime\prime}(\xi)}{2}. \tag{7.1.13}$$

By hypotheses, the term in the bracket tends to zero as $n \to \infty$ so taking the limit of both sides in (7.1.9) we have, by virtue of the preceding steps

$$\lim_{n \to \infty} t_i(n) = \emptyset_i(cy_i). \tag{7.1.14}$$

Since (7.1.14) holds for $i = 1,2,\cdots,m$ we conclude that

$$\lim_{n \to \infty} T^{(n)}(p + B^n y) = \emptyset(cy). \quad \square$$

If we had used the hypotheses for Theorem 7.1.2, we would clearly satisfy the hypotheses $a_{ij}(n)b_{j1}(n)b_{1k}(n) \overset{n}{\to} 0$, since that theorem gives this result for <u>all</u> possible arrangements of subscripts and is a stronger result than needed. If $m = 2$, similar comments also apply with regard to Corollary 7.1.1.

<u>Theorem 7.1.4</u>: If $m > 2$ and $T(x)$ satisfies (1), (2) and (3) where A satisfies the hypotheses of Theorem 7.1.2, then condition (ii) holds.

<u>Proof</u>: Let $\emptyset^*(x) = F(x)$ where F is of class C_1 on some neighborhood of p. Then after a componentwise argument we may write the resulting increment equation,

$$F(p + h) - F(p) = F'(p) \cdot h + Q \cdot h. \tag{7.1.15}$$

Now Q is an $m \times m$ matrix, each of whose elements is $o(||h||)$ where $||h||$ is the euclidean norm of h. Replacing h by $B^n y$ where $B^n y \overset{n}{\to} 0$ for each y, we have each element of $A^n Q(B^n y)$ as a finite sum of terms of form,

$$w(n) = a_{ij}(n) \cdot b_{rs}(n) \cdot o(||B^n y||) \cdot y_s. \tag{7.1.16}$$

Since $c|b_{kl}(n)| \geq ||B^n y||$ for some constant c and some subscripts k and l, we have by virtue of (7.1.16) that

$$|w(n)| \leq |a_{ij}(n)| \cdot |b_{rs}(n)| \cdot |b_{kl}(n)| \cdot \frac{o(||B^n y||)}{||B^n y||} \cdot cy_s .$$

But by hypothesis, $a_{ij}(n) \cdot b_{kl}(n) \cdot b_{rs}(n) \overset{n}{\to} 0$ for all choices of subscripts so $w(n) \overset{n}{\to} 0$. Therefore $A^n Q(B^n y) \overset{n}{\to} 0$ which is condition (ii). □

As a consequence of this argument we can state the corollary for the case m = 2.

<u>Corollary 7.1.2</u>: If m = 2 and T(x) satisfies (1), (2) and (3) where A satisfies the hypotheses of Corollary 7.1.1, then condition (ii) holds.

We shall provide two familiar examples of multivariate maps and obtain their limit maps under scaling and iteration.

<u>Example 7.1.1</u>: Let $T(x) = \begin{cases} t_1 = 1 - (1 - x)^{a_{11}}(1 - x_2)^{a_{12}} \\ t_2 = 1 - (1 - x)^{a_{21}}(1 - x_2)^{a_{22}} \end{cases}$ where $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$ satisfies the hypotheses of Theorem 7.1.3. Then if $\phi_1(u) = \phi_2(u) = 1 - \exp u$ and $\phi_1^*(u) = \phi_2^*(u) = \ln(1 - y)$ we have $\phi^{*\prime}(0) = -I$. By Theorem 7.1.3 with p = 0 we have

$$T^{(n)}(B^n y) \overset{n}{\to} \phi(-y) = \begin{bmatrix} 1 - e^{-y_1} \\ 1 - e^{-y_2} \end{bmatrix}.$$

Example 7.1.2: Let $T(x) = \frac{Ax}{cx + 1}$ satisfy (1), (2) and (3) where A is an $m \times m$ matrix, c is a $1 \times n$ vector and x is a $1 \times n$ vector. Assume the proper domain as noted in the discussion of linear fractional maps in Chapter III.

Setting $\emptyset(u) = \frac{u}{c(A - I)^{-1}u + 1}$ and $\emptyset^*(u) = \frac{u}{1 - c(A - I)^{-1}u}$, we can observe from computation that $\emptyset^{*\prime}(0) = I$.

$$T^{(n)}(B^n y) = \emptyset A^n \emptyset^*(B^n y) = \frac{A^n B^n y}{c(A - I)^{-1}(A^n - I)B^n y + 1}$$

$$= \frac{y}{c(A - I)^{-1}y - c(A - I)^{-1}B^n y + 1}.$$

Since $B^n \to 0$ we have the limit

$$T^{(n)}(B^n y) \overset{n}{\to} \frac{y}{c(A - I)^{-1}y + 1} = \emptyset(y).$$

In this example, conditions (i) and (ii) were met provided $B^n \to 0$.

B. Scaling with an Approximate Inverse

Suppose that A is singular or for some reason we do not use A^{-1} as the scaling matrix. Referring to Chapter VI we may employ the idea of an approximate inverse.

Let V be a nonsingular invariant subspace of A, denoted nisA. Let $B = A_V^{-1}$ be an approximate inverse of A with respect to V such that for any $x \in \mathcal{R}_m$,

$$A^k B^k(x) \overset{k}{\to} v_x, \; v_x \in V. \tag{7.2.1}$$

The previous assumptions (1), (2) and (3) need only be modified by replacing (2) with:

(2)' $T'(p) = A$ where B is an approximate inverse of A as in (7.2.1).

The sufficient conditions (i) and (ii) will remain unchanged. Following the steps leading to Equation (7.1.5) of the previous section we obtain analogously

$$\lim_{n \to \infty} T^{(n)}(p + B^n y) = \emptyset\{[\emptyset^{*\prime}(p)]v_y\}, \tag{7.2.2}$$

where $v_y \in V$.

This result indicates the possibility of obtaining a limit map under more general matrix scaling conditions than previously. Obviously, scaling with the inverse matrix (when it exists) is a special case. Other special cases are scaling with a conditional inverse or, if it exists, scaling with the dominant root.

<u>Example 7.2.1</u>: Consider $T(x) = \emptyset A \emptyset^{-}(x)$ where $\emptyset$ is a diagonal operator as in Theorem 7.1.3 and Corollary 7.1.1 with $m = 3$. Assume A is diagonalizable with characteristic roots $\lambda_1, \lambda_2, 0$ where $1 < |\lambda_2| < |\lambda_1| < |\lambda_2^2|$. Then $A = P[\mathrm{diag}(\lambda_1, \lambda_2, 0)]P^{-1}$ where we let P_1, P_2, P_3 be column vectors of P and Q_1, Q_2, Q_3 be row vectors of $P^{-1} = Q$.

Let $B = P[\mathrm{diag}(\lambda_1^{-1}, \lambda_2^{-1}, 0)]P^{-1}$ be an approximate inverse of A relative to the nisA spanned by P_1, P_2. We have $A^n = \lambda_1^n P_1 Q_1 + \lambda_2^n P_2 Q_2$ and $B^n = \lambda_1^{-n} P_1 Q + \lambda_2^{-n} P_2 Q_2$. Then for the ij^{th} element of A^n, $a_{ij}(n)$, we see that $|a_{ij}(n)| = |c_1 \lambda_1^n + c_2 \lambda_2^n| \le |c_3 \lambda_1^n|$ for some constants c_1, c_2, c_3.

Also, $(B_j^{(n)}y)^2 = \lambda_1^{-2n}(d_1y)^2 + (\lambda_1\lambda_2)^{-n}(d_2y)^2 + \lambda_2^{-2n}(d_3y)^2 \le |\lambda_2^{-2n}|(d_4y)^2$ for some constants d_1,d_2,d_3,d_4. Then, $|a_{ij}(n)|(B^{(n)}(y))^2 \le \left|\frac{\lambda_1}{\lambda_2^2}\right|^n \cdot k \overset{n}{\to} 0$, where k is some constant. This is sufficient to cause the right-hand sum of (7.1.11) to tend to zero, as desired. Clearly, the left-hand sum of (7.1.11), $\sum_{j=1}^{3} a_{ij}(n) \cdot B_j^{(n)}y = v_i$, is the i^{th} component of a vector where $v_3 = 0$. Then for any $y \varepsilon R_3$, $A^nB^ny = v_y = \begin{bmatrix} v_1 \\ v_2 \\ 0 \end{bmatrix}$, a vector in the nisA spanned by P_1,P_2.

If we had chosen $B = P[\mathrm{diag}(\lambda_1^{-1},0,0)]P^{-1}$, we obtain the result analogous to scaling with the dominant root.

C. Multitype Branching Process Example

Harris (1951), gives a discussion of the multitype Galton-Watson process wherein he hints the possibility of using the moment matrix for scaling. Karlin and McGregor (1970), pursue this approach in a certain example. We shall illustrate by observing a bivariate example which could readily generalize to R_n.

Suppose there are two types of particles, type x and type y. Let X_n and Y_n represent the number of particles of each type respectively at the n^{th} generation. Let p_{ij} be the probability that a particle of type x will have i offspring of type x and j offspring of type y. Let q_{ij} be the probability that a particle of type y will have i offspring of type x and j offspring of type y.

Let $Z_n = \begin{bmatrix} X_n \\ Y_n \end{bmatrix}$, $s = \begin{bmatrix} s_1 \\ s_2 \end{bmatrix}$ and $e^s = \begin{bmatrix} e^{s_1} \\ e^{s_2} \end{bmatrix}$ where $0 < s_i \leq 1$; $i = 1,2$.

Harris (1963), page 36, gives a theorem stating that the multivariate probability generating function for Z_n is, as in R_1, the n^{th} iterate of the probability generating function of Z_1. We shall indicate notationally this iteration of the bivariate probability generating function by first letting the bivariate probability generating function of Z_1 be

$$F(s) = \begin{cases} f_1(s) = \sum_{j=o}^{\infty} \sum_{i=o}^{\infty} p_{ij} s_1^i s_2^j \\ \\ f_2(s) = \sum_{j=o}^{\infty} \sum_{i=o}^{\infty} q_{ij} s_1^i s_2^j . \end{cases} \tag{7.3.1}$$

The probability generating function of Z_n is then denoted

$$F^{(n)}(s) = F(F^{(n-1)}(s)) = \begin{cases} f_1(F^{(n-1)}(s)) \\ \\ f_2(F^{(n-1)}(s)) \end{cases} . \tag{7.3.2}$$

Let $Z_o = [{1 \atop 0}]$ be the starting vector and define for $s_* = [{1 \atop 1}]$,

$$A = F'(s_*^-) = \begin{bmatrix} \frac{\partial f_1}{\partial s_1} & \frac{\partial f_1}{\partial s_2} \\ \\ \frac{\partial f_2}{\partial s_1} & \frac{\partial f_2}{\partial s_2} \end{bmatrix}_{(s_*)} \quad \text{to be the first moment matrix}$$

(where the derivatives are left partial derivatives). Then $E[Z_1] = AZ_o$ is the mean vector for Z_1. The second moment matrix would be given, conditional to $Z_o = [{1 \atop 0}]$, by $[F_1''(s_*)] - [\mathrm{diag}(a_{11}, a_{21})]$ where $F_1''(s_*)$ is

$$F_1''(s_*^-) = \begin{bmatrix} \dfrac{\partial^2 f_1}{\partial s_1 \partial s_1} & \dfrac{\partial^2 f_1}{\partial s_1 \partial s_2} \\ \\ \dfrac{\partial^2 f_1}{\partial s_2 \partial s_1} & \dfrac{\partial^2 f_1}{\partial s_2 \partial s_2} \end{bmatrix}_{(s_*)} .$$

The R_2 analogy to Var Z_1 is the covariance matrix of Z_1 given Z_0,

$$[F_1''(s_*)] - [\mathrm{diag}(a_{11}, a_{21})] + [AZ_0][AZ_0]^T. \qquad (7.3.3)$$

(T indicates the transpose of the column vector.)

Basic assumptions are:

(1) First and second moment matrices are finite.

(2) $\{Z_n\}$ is a Markov chain.

(3) The absolute values of the characteristic roots of A exceed unity.

(The third assumption may be modified by appealing to the approximate inverse idea.)

It seems possible to follow a completely analogous argument based on martingales as in Chapter IV where if $W_n = B^n Z_n$, B being A^{-1}, we have $W_n \overset{n}{\rightarrow} W$ almost surely and $E[W_n] \overset{n}{\rightarrow} E[W] = IZ_0 = Z_0$.

The symbolic notation works equally well for the bivariate moment generating function,

$$\phi_{Z_n}(s) = \begin{cases} \phi_n(s) \\ \\ \psi_n(s) \end{cases} = \begin{cases} f_1(F^{(n-1)}(e^s)) \\ \\ f_2(F^{(n-1)}(e^s)) \end{cases} = F^{(n)}(e^s). \qquad (7.3.4)$$

It may be seen that the usual computational properties of moment generating functions hold in this symbolic notation by examining it componentwise.

Then we may proceed,

$$\emptyset_{W_n}(s) = \emptyset_{B^n Z_n}(s) = \emptyset_{Z_n}(B^n s) = F^{(n)}(e^{B^n s}).$$

Therefore, we have

$$\emptyset_{W_n}(As) = F(F^{(n-1)}(e^{B^{n-1}s})). \qquad (7.3.5)$$

By preceding remarks, note that $\emptyset_{W_n}(s) \overset{n}{\to} \emptyset_W(s)$.

Taking limits in (7.3.5) after letting $\emptyset_W = \emptyset$ we have

$$\emptyset(As) = F(\emptyset(s)); \ -\infty < s_i \leq 0, \ i = 1,2. \qquad (7.3.6)$$

Setting $t = \emptyset(s)$ where $0 < t_i \leq 1$, $i = 1,2$ gives

$$F(t) = \emptyset(A\emptyset^{-}(t)). \qquad (7.3.7)$$

Now, s_* is a fixed point of $F(t)$ so centering at s_* and scaling by B^n we have after iteration of (7.3.7) and setting $t = s_* + B^n s$

$$F^{(n)}(s_* + B^n s) = \emptyset(A^n \emptyset^{-}(s_* + B^n s). \qquad (7.3.8)$$

Proceeding as in the first section of this chapter, since $\emptyset'(\underline{0}^{-}) = I$ the commutativity requirement (i) is satisfied. We must postulate that $A^n R(B^n s) \overset{n}{\to} 0$, to satisfy condition (ii) and hence obtain the result,

$$F^{(n)}(s_* + B^n s) \overset{n}{\to} \emptyset(s); \ -\infty < s_i \leq 1, \ i = 1,2. \qquad (7.3.9)$$

As previously noted, we may relax requirement (3) and employ an approximate inverse, however in the bivariate case this would be the trivial case of scaling by the dominant root, assuming it exists.

VIII. BIVARIATE BOUND MAPS

Before discussing possibilities for bivariate bound maps we shall sketch the bivariate analog for the existence of a limit function under scaling and iteration as exhibited by Thomas and David (1968). Since the function (the maximin function) being iterated was of a special nature one would expect to have rather restrictive conditions for an analogous mapping in R_2.

Suppose T(x) is a differentiable mapping from D onto D for $D \subset R_2$, where T is componentwise (i) convex, (ii) nondecreasing and (iii) bounded above. Let p be a fixed vector of T and let $T'(p) = A$ be a nonsingular matrix with characteristic roots λ_1 and λ_2 such that $\min(|\lambda_1|,|\lambda_2|) > 1$.

Since T is componentwise convex, by Definition 2.1.5 we have the vector of componentwise tangent planes given by

$$T(x) \geq T(p) + A(x - p);\ x \in D. \tag{8.1.1}$$

If $B = A^{-1}$, B^n converges by Theorem 7.1.1. Then for n sufficiently large we have $p + B^n z \in D$ for any $z \in R_2$. Hence, setting $x = p + B^n z$ in (8.1.1) we obtain

$$T(p + B^n z) \geq p + AB^n z = p + B^{n-1}z;\ z \in R_2. \tag{8.1.2}$$

Since T is componentwise nondecreasing, so is $T^{(n-1)}$ and taking $T^{(n-1)}$ of both sides in (8.1.2) for n sufficiently large gives

$$T^{(n)}(p + B^n z) \geq T^{(n-1)}(p + B^{n-1}z);\ z \in R_2. \tag{8.1.3}$$

By virtue of (8.1.3) we say that $\{T^{(n)}(p + B^n z)\}$ constitutes an eventually componentwise nondecreasing sequence of mappings. Since T is bounded above, componentwise, a limit map $L_T(z)$ exists as $n \to \infty$.

Clearly, by changing (i), (ii) and (iii) respectively to (i)' concave, (ii)' nonincreasing and (iii)' bounded below, we could employ the same reasoning and obtain existence of $L_T(z)$.

In attempting to continue the bivariate analog to bound maps for iterates of T we encounter difficulties complicated by the lack of a definition for "boundedness", so we make the following definition.

Definition 8.1.1: If R, S and T are mappings from R_n to R_n, R(x) and S(x) shall be said to bound T(x) on D if each component of T(x) is between the corresponding components of R(x) and S(x) for $x \in D$.

("Between" is used in the sense that a is between b and c if $b \leq a \leq c$ or $c \leq a \leq b$.)

Another complication in extending bound function concepts in R_1 to bound map concepts in R_2 is that the restrictions on T become quite severe. The author has been unable to produce examples of iterated maps bounded in this sense, however, we shall assume that for some sequence $T_n(x)$, not necessarily iterates of T, that the bound maps to be used will satisfy the definition for each n.

A reasonable approach is that the bound maps have a common fixed point of interest and are similar in the sense we defined, that is, have the same core matrix.

We previously exploited the technique of translating the fixed point to the origin, so without loss of generality let us consider p = 0 as the common fixed vector for R, S and T. We shall consider R(x) and S(x) as bivariate analogs of λ and μ used in Chapter IV.

$$R(x) = \begin{cases} r_1 = u_1 - u_1\left(\dfrac{u_1 - x_1}{u_1}\right)^{a_{11}}\left(\dfrac{u_2 - x_2}{u_2}\right)^{a_{12}} \\ r_2 = u_2 - u_2\left(\dfrac{u_1 - x_1}{u_1}\right)^{a_{21}}\left(\dfrac{u_2 - x_2}{u_2}\right)^{a_{22}} \end{cases} \tag{8.1.4}$$

where $u_i > 0$ and $x_i \leq u_i$ for $i = 1,2$.

$$S(x) = \begin{cases} s_1 = v_1 - v_1\left(\dfrac{v_1 - x_1}{v_1}\right)^{a_{11}}\left(\dfrac{v_2 - x_2}{v_2}\right)^{a_{12}} \\ s_2 = v_2 - v_2\left(\dfrac{v_1 - x_1}{v_1}\right)^{a_{21}}\left(\dfrac{v_2 - x_2}{v_2}\right)^{a_{22}} \end{cases} \tag{8.1.5}$$

where $v_i < 0$ and $x_i \geq v_i$ for $i = 1,2$.

When these are related to Theorem 7.1.3, Corollary 7.1.1 and Example 7.1.1 we see that $R(x)$ and $S(x)$ are easily iterated similar maps having edge maps expressed in the form of "diagonal $\emptyset$" type matrix operators. In this form we have the following equations.

$$R(x) = \emptyset A \emptyset^{-}(x) = \begin{bmatrix} \emptyset_1 & 0 \\ \emptyset & \emptyset_2 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} \emptyset_1^{-} & 0 \\ 0 & \emptyset_2^{-} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \tag{8.1.6}$$

where $\emptyset_i(t) = u_i(1 - \exp t)$ and $\emptyset_i^{-}(t) = \ln\left(\dfrac{u_i - t}{u_i}\right)$; $i = 1,2$.

$$S(x) = \theta A \theta^{-}(x) = \begin{bmatrix} \theta_1 & 0 \\ 0 & \theta_2 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} \theta_1^{-} & 0 \\ 0 & \theta_2^{-} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \tag{8.1.7}$$

where $\theta_i(t) = v_i(1 - \exp t)$ and $\theta_i^{-}(t) = \ln\left(\dfrac{v_i - t}{v_i}\right)$; $i = 1,2$.

Let A have characteristic roots λ_1, λ_2 with $|\lambda_i| > 1$ for $i = 1,2$. Also, let A be a nonnegative matrix. An example is

$$A = \begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix} \text{ where } \lambda_1 = 2, \lambda_2 = 3, B = \begin{bmatrix} \frac{1}{2} & -\frac{1}{6} \\ 0 & \frac{1}{3} \end{bmatrix}.$$

(Note that $\lambda_2 < \lambda_1^2$, the hypothesis of Corollary 7.1.1.)

Denote $u = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$, $v = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$ and $0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$.

It is easy to see that $R(u) = u$, $S(v) = v$ and $R(0) = S(0) = 0$. Since we wish $R'(0) = S'(0) = A$, by Corollary 3.1.2 it is sufficient to have $\emptyset'(0)$ and $\theta'(0)$ commute with A. One way to accomplish this is to let $u_1 = u_2$ and $v_1 = v_2$ which gives $\emptyset'(0) = u_1 I$ and $\theta'(0) = v_1 I$. This means that opposite vertices of the rectangular domain for $T(x)$ will be on the equiangular line with the origin always in the rectangle.

To illustrate the situation without too much loss of generality but considerable simplification notationally, let $u = \underline{e} = [{1 \atop 1}]$ and $v = -\underline{e} = [{-1 \atop -1}]$ be the opposite vertices of the square domain. Then (8.1.4) and (8.1.5) become

$$R(x) = \begin{cases} 1 - (1 - x_1)^{a_{11}}(1 - x_2)^{a_{12}} \\ 1 - (1 - x_1)^{a_{21}}(1 - x_2)^{a_{22}} \end{cases} ; x \in D_R = \{x : x \le \underline{e}\}, \tag{8.1.8}$$

$$S(x) = \begin{cases} -1 + (1 + x_1)^{a_{11}}(1 + x_2)^{a_{12}} \\ -1 + (1 + x_1)^{a_{21}}(1 + x_2)^{a_{22}} \end{cases} ; x \in D_S = \{x : x \ge -\underline{e}\}. \tag{8.1.9}$$

Thinking of these in the operator matrix form of (8.1.6) and (8.1.7) the edge map components are $\emptyset_i(t) = 1 - \exp t$, $\emptyset_i^-(t) = \ln(1 - t)$ and $\theta_i(t) = -1 + \exp t$, $\theta_i^-(t) = \ln(1 + t)$, $i = 1,2$.

This gives us $\emptyset'(0) = I$ and $\theta'(0) = -I$. The domain of the bounded map is then $D_R \cap D_S = D$, the square mentioned earlier. $R(x)$ will be considered a lower bound map and $S(x)$ will be considered an upper bound map.

<u>Theorem 8.1.1</u>: $R(x)$ and $S(x)$ are componentwise nondecreasing on D_R and D_S respectively.

<u>Proof</u>: Let $x' = \begin{bmatrix} x_1' \\ x_2' \end{bmatrix}$ and $x'' = \begin{bmatrix} x'' \\ x_2'' \end{bmatrix}$ be in D_R such that $x' \leq x''$.

Looking at the first component only, $x_1' \leq x_1''$ and $x_2' \leq x_2''$ imply $(1 - x_1') \geq (1 - x_1'')$ and $(1 - x_2') \geq (1 - x_2'')$. Then since A is nonnegative, $(1 - x_1')^{a_{11}}(1 - x_2')^{a_{12}} \geq (1 - x_1'')^{a_{11}}(1 - x_2'')^{a_{12}}$. Multiplying by -1 and adding 1 gives the first component of R and the inequality $r_1(x') \leq r_1(x'')$. Doing the same for $r_2(x)$, we have the result for $R(x)$. Looking at the first component of $S(x)$ it is clear that $-1 + (1 + x_1')^{a_{11}} \times (1 + x_2')^{a_{12}} \leq -1 + (1 + x_1'')^{a_{11}}(1 + x_2'')^{a_{12}}$, and doing the same for the second component we have the result for $S(x)$. □

Denote $A^n = \begin{bmatrix} a_{11}(n) & a_{12}(n) \\ a_{21}(n) & a_{22}(n) \end{bmatrix}$ and where $B = A^{-1}$,

$B^n = \begin{bmatrix} b_{11}(n) & b_{12}(n) \\ b_{21}(n) & b_{22}(n) \end{bmatrix}$. By the nonnegativity of A and the characteristic root requirements, $a_{11}(n) \overset{n}{\to} +\infty$ or $a_{12}(n) \overset{n}{\to} +\infty$ (or both) and $a_{21}(n) \overset{n}{\to} +\infty$ or $a_{22}(n) \overset{n}{\to} +\infty$ (or both). From Theorem 7.1.1, $b_{ij}(n) \overset{n}{\to} 0$ for all i,j.

We then see that for $x_1, x_2 < 0$,

$$R^{(n)}(x) = \begin{cases} 1 - (1 - x_1)^{a_{11}(n)}(1 - x_2)^{a_{12}(n)} \\ \\ 1 - (1 - x_1)^{a_{21}(n)}(1 - x_2)^{a_{22}(n)} \end{cases} \xrightarrow{n} \begin{bmatrix} -\infty \\ -\infty \end{bmatrix}$$

and for $x_1, x_2 > 0$,

$$S^{(n)}(x) = \begin{cases} -1 + (1 + x_1)^{a_{11}(n)}(1 + x_2)^{a_{12}(n)} \\ \\ -1 + (1 + x_1)^{a_{21}(n)}(1 - x_2)^{a_{22}(n)} \end{cases} \xrightarrow{n} \begin{bmatrix} +\infty \\ +\infty \end{bmatrix}.$$

For $x_1, x_2 < 0$ the componentwise lower bound is $-\underline{e}$ and for $x_1, x_2 > 0$ the componentwise upper bound is $\underline{e}$.

We are interested in determining what conditions on z are necessary for $R^{(n)}(B^n z)$ and $S^{(n)}(B^n z)$ to be on D.

<u>Theorem 8.1.2</u>: If $-B^n q \leq B^n z \leq B^n q$ for $n = 0,1,2,\cdots$ where $q = \begin{bmatrix} \ln 2 \\ \ln 2 \end{bmatrix}$, then $R^{(n)}(B^n z)$ and $S^{(n)}(B^n z)$ are on D.

<u>Proof</u>: For any real number $w > 0$, we have $1 - w \leq -\ln w$ and $w - 1 \geq \ln w$. Let $B_1(n)$ and $B_2(n)$ be the row sums of B^n. We obtain the following:

$$R^{-(n)}(-\underline{e}) = \begin{cases} 1 - (2)^{B_1(n)} \\ \\ 1 - (2)^{B_2(n)} \end{cases} \leq \begin{cases} -B_1(n)(\ln 2) \\ \\ -B_2(n)(\ln 2) \end{cases} = -B^n q. \qquad (8.1.10)$$

$$S^{-(n)}(\underline{e}) = \begin{cases} -1 + (2)^{B_1(n)} \\ \\ -1 + (2)^{B_2(n)} \end{cases} \geq \begin{cases} B_1(n)(\ln 2) \\ \\ B_2(n)(\ln 2) \end{cases} = B^n q. \tag{8.1.11}$$

By (8.1.10), (8.1.11) and the hypotheses of the theorem we have

$$R^{-(n)}(-\underline{e}) \leq -B^n q \leq B^n z \leq B^n q \leq S^{-(n)}(\underline{e}). \tag{8.1.12}$$

From the componentwise monotonicity of R and hence $R^{(n)}$ we have upon taking $R^{(n)}$ of the left and center terms of (8.1.12),

$$-\underline{e} \leq R^{(n)}(B^n z). \tag{8.1.13}$$

Similarly, taking $S^{(n)}$ of the center and right terms of (8.1.12), we have

$$S^{(n)}(B^n z) \leq \underline{e}. \tag{8.1.14}$$

By definition, $R^{(n)}(\cdot) \leq \underline{e}$ and $S^{(n)}(\cdot) \geq -\underline{e}$ so from (8.1.13) and (8.1.14) we get the result,

$$-\underline{e} \leq R^{(n)}(B^n z) \leq \underline{e}$$

and $-\underline{e} \leq S^{(n)}(B^n z) \leq \underline{e}$. $\square$

It is not true, in general, that an inequality is preserved when multiplying by a matrix that is not nonnegative. Therefore, the condition $-q < z < q$ is not sufficient for the hypotheses of the theorem. If we suppose B to be nonnegative, then A will not necessarily be so and we may lose the componentwise monotonicity for R(x) and S(x). It is necessary that $-q \leq p \leq q$, since that is the case when n = 0 in the hypotheses.

Let A satisfy the hypotheses of Theorem 7.1.3. Then since $\emptyset'(0) = I$ and $\theta'(0) = -I$, we have the following limit maps:

$$\lim_{n \to \infty} R^{(n)}(B^n z) = \emptyset(z) = \begin{cases} 1 - e^{z_1} \\ 1 - e^{z_2} \end{cases} ; \; z \in \mathcal{R}_2, \tag{8.1.15}$$

$$\lim_{n \to \infty} S^{(n)}(B^n z) = \theta(-z) = \begin{cases} -1 + e^{-z_1} \\ -1 + e^{-z_2} \end{cases} ; \; z \in \mathcal{R}_2. \tag{8.1.16}$$

Setting z = q and -q respectively in (8.1.15) and (8.1.16) gives, as expected, the opposite vertices of D.

$$\lim_{n \to \infty} R^{(n)}(B^n q) = \begin{bmatrix} 1 - e^{\ln 2} \\ 1 - e^{\ln 2} \end{bmatrix} = \begin{bmatrix} -1 \\ -1 \end{bmatrix} = -\underline{e},$$

$$\lim_{n \to \infty} S^{(n)}(-B^n q) = \begin{bmatrix} -1 + e^{\ln 2} \\ -1 + e^{\ln 2} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \underline{e}.$$

The preceding discussion is intended to show the possible potential of these particular types of bound maps which when scaled, could "control" map iterates by keeping them on D for z in some domain. The nature of maps to be bounded in this sense remains an open question. An intuitive approach would be to consider a map formed by some sort of averaging of the corresponding components of R and S.

IX. LITERATURE CITED

Buck, R. C. (1965). Advanced Calculus. 2nd edition. McGraw-Hill, New York.

Doob, J. L. (1953). Stochastic Processes. Wiley, New York.

Feller, William (1968). An Introduction to Probability Theory and Its Applications. 3rd edition. Vol. 1. Wiley, New York.

Fisz, Marek (1963). Probability Theory and Mathematical Statistics. 3rd edition. Wiley, New York.

Fleming, W. H. (1965). Functions of Several Variables. Addison-Wesley, Dallas.

Harris, T. E. (1951). Some mathematical models for branching processes. Berkeley Symposium in Probability 2:305-328.

Harris, T. E. (1963). The Theory of Branching Processes. Springer-Verlag, Berlin.

Karlin, S. and McGregor J. (1970). Iteration of analytic functions of several variables. In R. C. Gunning, ed. Papers in Honor of Salomon Bochner. Pp. 81-92. Princeton University Press, Princeton, N. J.

Koenigs, M. G. (1884). Recherches sur les integrales of certaines equations fonctionelles. Annales Scientifiques de l'Ecole Normale Superieure. 1, Supplement: S.2-S.41.

Rudin, Walter (1953). Principles of Mathematical Analysis. McGraw-Hill, New York.

Schroeder, Ernst (1871). Ueber iterirte functionen. Mathematische Annalen 3:296-322.

Schubert, C. F. (1964). Solution of a generalized Schroeder equation in two variables. Journal of the Australian Mathematical Society 4:410-417.

Szekeres, G. (1958). Regular iteration of real and complex functions. Acta Mathematica 100:203-258.

Thomas, D. R. (1965). Asymptotic value distributions for certain $2 \times n$ games and n-stage games of perfect information. Ph.D. dissertation, Iowa State University.

Thomas, D. R. and David, H. T. (1968). Game value distributions I. Annals of Mathematical Statistics 38:242-250.

Titchmarsh, E. C. (1939). The Theory of Functions. 2nd edition. Oxford University Press, London.

Tucker, H. G. (1967). A Graduate Course in Probability. Academic Press, New York.

Varga, R. (1962). Matrix Iterative Analysis. Prentice-Hall, Englewood Cliffs, New Jersey.

X. ACKNOWLEDGEMENTS

The author is indebted to Dr. R. C. Bzoch whose efforts helped him gain the opportunity for further graduate study. The encouragement of colleagues at the University of North Dakota and the cooperation from the statistics faculty at Iowa State University have been appreciated. Dr. T. A. Bancroft's kindly advice, especially during the first year, was very beneficial.

To his wife and children, who tolerated his idiosyncrasies during the past four years, the author expresses his deepest appreciation for their consideration and understanding.

To Dr. H. T. David who patiently and unselfishly provided his time and resources along with his ability to instill and sustain a certain amount of self-confidence so necessary for such an endeavor, the author can only offer an inadequate but sincere "thank you".

XI. APPENDIX A

The maximin function $\phi(t) = [1 - (1 - t)^2]^2$ and the minimax function $\psi(t) = 1 - (1 - t^2)^2$ were developed from a probabilistic approach in Chapter IV. A somewhat different approach to the "construction" of these functions is to think of them as being generated by second iterates of quadratic functions.

If $g(x) = (1 - x)^2$, then $g(g(x)) = \phi(x)$ and if $h(x) = 1 - x^2$, then $h(h(x)) = \psi(x)$. Notice that $g(0) = 1$, $g(1) = 0$, $h(0) = 1$ and $h(1) = 0$ so $g(x)$ and $h(x)$ are parabolas both having intercepts at the points $(0,1)$ and $(1,0)$. Furthermore, any fixed point of $g(x)$ or $h(x)$ will have to be fixed points respectively of $\phi(x)$ or $\psi(x)$. Since $g(x) = x$ at $x = \frac{1}{2}(3 \pm \sqrt{5})$ and $h(x) = x$ at $x = \frac{1}{2}(-1 \pm \sqrt{5})$ we see that $\phi(x)$ has a fixed point on the interval $0 < x < 1$ at $x_1 = \frac{1}{2}(3 - \sqrt{5})$ and $\psi(x)$ has a fixed point on the interval $0 < x < 1$ at $x_2 = \frac{1}{2}(-1 + \sqrt{5})$. It can easily be checked that these are unique fixed points on the given interval for $\phi(x)$ and $\psi(x)$ respectively.

We find that $x_1 + x_2 = 1$ establishes a sort of "complementary" relation between the fixed points of $\phi(x)$ and $\psi(x)$ on $0 < x < 1$. It is further easily established that $\phi'(x_1) = \psi'(x_2) = b > 1$ for some number b. A more general result is contained in the following theorem.

<u>Theorem A.1</u>: If $f(x) = cx^2 - (c + 1)x + 1$, let $F(x;c) = f^{(2)}(x)$. Then, if a is a fixed point of $F(x;c)$, $1 - a$ is a fixed point of $F(x;-c)$ where $-1 \leq c \leq 1$, $c \neq 0$. Furthermore, $F'(a;c) = F'(1 - a;-c) = b > 1$.

Proof: Observe that if we begin with $f(x) = cx^2 + dx + e$, then $f(0) = 1$ and $f(1) = 0$ causes $e = 1$ and $d + c = -1$. Letting $d = -(c + 1)$ and $e = 1$ gives $f(x)$ as in the theorem. The fixed points of $F(x;c)$ are roots to the equation $F(x;c) - x = 0$ which becomes

$$c^3x^4 - 2c^2(c + 1)x^3 + c^3(c + 3)x^2 + (1 - c^2)x - x = 0$$

or $$c^2x(x - 1)[cx^2 - (c + 2)x + 1] = c^2x(x - 1) \cdot f(x) = 0.$$

Therefore, fixed points of $F(x;c)$ are found to be 0, 1 and $[(c + 2) \pm \sqrt{c^2 + 4}]/2c$. The unique fixed point on $0 < x < 1$ when $0 < c \leq 1$ is $[(c + 2) - \sqrt{c^2 + 4}]/2c$. Similar computations show that for $-1 \leq c < 0$ the unique fixed point on $0 < x < 1$ is $[(c + 2) + \sqrt{c^2 + 4}]/2c$.

Hence, for $F(x;c)$ where $c > 0$, we have $F(x;c) = x$ at $a = [(c + 2) - \sqrt{c^2 + 4}]/2c$ and $F(x;-c) = x$ at $1 - a = [(-c + 2) + \sqrt{c^2 + 4}] /(-2c)$ where both a and $1 - a$ are on the interval $0 < x < 1$. If c is restricted as in the theorem, further computations will show that the derivative relation holds, that is, $F'(a;c) = F'(1 - a;-c) = b$ where $1 < b < +\infty$. $\square$

XII. APPENDIX B

To illustrate why the conditions of Corollary 7.1.1 are considered as minimally nontrivial we shall exhibit the details of one of the sixteen possible cases in the $m = 2$ situation.

Theorem B.1: Let A be a 2×2 matrix such that $A^{-1} = B$ and if λ_1, λ_2 are real eigenvalues of A, let $1 < |\lambda_1| \leq |\lambda_2|$. If $P = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is such that $P^{-1}AP = J$, the Jordan canonical form, then letting $a_{ij}(n)$ and $b_{ij}(n)$ be the ij^{th} elements of A^n and B^n respectively we have $a_{11}(n)b_{11}^2(n) \overset{n}{\to} 0$ under the following conditions:

(1) $\lambda_1 = \lambda_2$ and $|\lambda_2| < \lambda_1^2$

or

(2) $\lambda_1 \neq \lambda_2$ and any one of the following holds:

(i) $|\lambda_2| < \lambda_1^2$.

(ii) $b[d + (\lambda_1 - \lambda_2)c] = 0$.

(iii) $d^2[b^2 d + 2ab(\lambda_1 - \lambda_2) + a^2(\lambda_1 - \lambda_2)^2] = 0$.

Proof: If $\lambda_1 = \lambda_2 = \lambda$; $A^n = PJ^nP^{-1} = P \begin{bmatrix} \lambda^n & 0 \\ n\lambda^{n-1} & \lambda^n \end{bmatrix} P^{-1}$

and $B^n = P \begin{bmatrix} \lambda^{-n} & 0 \\ -\dfrac{n}{\lambda^{n+1}} & \lambda^{-n} \end{bmatrix} P^{-1}$ where $P = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$,

then $|P|a_{11}(n) = [\lambda^n(|P| + \frac{nbd}{\lambda})]$

and $|P|^2 b_{11}^2(n) = [\lambda^{-n}(|P| - \frac{nbd}{\lambda})]^2$. Then we have

$$|P|^3 a_{11}(n) b_{11}^2(n) = \lambda^{-n}(|P| + \frac{nbd}{\lambda})(|P| - \frac{nbd}{\lambda})^2$$

which is of order of magnitude $\frac{n^3}{\lambda^n} \overset{n}{\to} 0$ since $|\lambda| > 1$.

Clearly $|\lambda_1| = |\lambda_2| > 1$ implies that $|\lambda_2| < \lambda_1^2$ and (1) is proved.

If $\lambda_1 \neq \lambda_2$, we have for (2)

$$|P| a_{11}(n) = [ad\lambda_1^n + bd\left(\frac{\lambda_1^n - \lambda_2^n}{\lambda_1 - \lambda_2}\right) - bc\lambda_2^n] \tag{B.1}$$

$$|P|^2 b_{11}^2(n) = [ad\lambda_1^{-n} + bd\left(\frac{\lambda_1^{-n} - \lambda_2^{-n}}{\lambda_1 - \lambda_2}\right) - bc\lambda_2^{-n}]^2. \tag{B.2}$$

Multiplying (B.1) times (B.2) gives

$$\begin{aligned} |P|^3 a_{11}(n) b_{11}^2(n) &= [ad\lambda_1^n + bd\left(\frac{\lambda_1^n - \lambda_2^n}{\lambda_1 - \lambda_2}\right) - bc\lambda_2^n] \\ &\times [a^2d^2\lambda_1^{-2n} + b^2d^2\left(\frac{\lambda_1^{-n} - \lambda_2^{-n}}{\lambda_1 - \lambda_2}\right) + b^2c^2\lambda_2^{-2n} \qquad (B.3) \\ &+ 2abd^2\lambda_1^{-n}\left(\frac{\lambda_1^{-n} - \lambda_2^{-n}}{\lambda_1 - \lambda_2}\right) - 2abcd\lambda_1^{-n}\lambda_2^{-n} - 2b^2cd\left(\frac{\lambda_1^{-n} - \lambda_2^{-n}}{\lambda_1 - \lambda_2}\right)\lambda_2^{-n}]. \end{aligned}$$

Since $|\lambda_1| < |\lambda_2|$, when the expression (B.3) is multiplied out, all terms in the first factor with λ_1^n and those in the second factor with λ_2^{-n} clearly will tend to zero. The terms of the second factor with λ_2^{-2n} will tend to zero regardless of the first factor. Calling these o(1) gives

$$|P|^3 a_{11}(n) b_{11}^2(n) = \left[\frac{-bd - (\lambda_1 - \lambda_2)bc}{\lambda_1 - \lambda_2}\right] \times$$

$$\left[\frac{b^2d^3 + 2abd^2(\lambda_1 - \lambda_2) + a^2d^2(\lambda_1 - \lambda_2)^2}{(\lambda_1 - \lambda_2)^2}\right]\left(\frac{\lambda_2}{\lambda_1^2}\right)^n + o(1).$$

Then $a_{11}(n)b_{11}^2(n) \overset{n}{\to} 0$ if any one of the following hold:

(i) $|\lambda_2| < \lambda_1^2$.

(ii) $b[d + (\lambda_1 - \lambda_2)c] = 0$.

(iii) $d^2[b^2d + 2ab(\lambda_1 - \lambda_2) + a^2(\lambda_1 - \lambda_2)^2] = 0$. □

Since (ii) and (iii) are very specialized situations we say (i) represents a minimally nontrivial sufficient condition.

It could be seen that the other 15 possible products in $a_{ij}(n)b_{jk}(n)b_{rs}(n)$ for any subscript would follow a similar argument.

To see why $|\lambda_2| = \lambda_1^2$ will not suffice, we have only to observe a counterexample:

Set $|\lambda_2| = \lambda_1^2$, say $\lambda_2 = 4$, $\lambda_1 = 2$. Then if $a = b = c = 2$ and $d = 1$ we have neither (ii) nor (iii) holding so the matrix for the counterexample is

$$A = \begin{bmatrix} 2 & 2 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} 2 & 2 \\ 2 & 1 \end{bmatrix}^{-1} . \quad \square$$

The same analysis as in Theorem B.1 holds for λ_1, λ_2 complex where absolute values are used throughout, i.e., $1 < |\lambda_1| < |\lambda_2|^2$ are the minimally nontrivial sufficient conditions as in Corollary 7.1.1.